

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 19, 2003, 18:01:21 : Search time 6.51797 Seconds
(without alignments)
6329.459 Million cell updates/sec

Title: US-09-997-610-1_COPY_56_136

Perfect score: 81
Sequence: 1 gacccacagcaccacccagc.....cacaatatacagagaata 81

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 424239 seqs, 25461826 residues

Total number of hits satisfying chosen parameters: 848478

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA.*
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq.*
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq.*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq.*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq.*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq.*
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq.*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq.*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq.*
9: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq.*
10: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq.*
11: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq.*
12: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq.*
13: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq.*
14: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	81	100.0	1381	9 US-09-997-610-1	Sequence 1, Appl1
2	81	100.0	1731	9 US-09-997-610-5	Sequence 5, Appl1
3	57	70.4	1377	9 US-09-997-610-3	Sequence 3, Appl1
4	57	70.4	1731	9 US-09-997-610-7	Sequence 7, Appl1
5	32.4	40.0	392	10 US-09-960-352-1786	Sequence 1786, Ap
6	32.4	40.0	415	10 US-09-960-352-9137	Sequence 9137, Ap
7	30.4	37.5	549	10 US-09-923-779-49	Sequence 49, Appl
8	30.4	37.5	3394	10 US-09-880-107-2178	Sequence 2178, Ap
9	30	37.0	2686	9 US-09-764-868-48	Sequence 48, Appl
10	30	37.0	3580	10 US-09-799-799-1	Sequence 1, Appl1
11	29.8	36.8	561	10 US-09-923-779-47	Sequence 47, Appl
12	29.8	36.8	617	10 US-09-923-779-51	Sequence 51, Appl
13	29.8	36.8	648	10 US-09-923-779-56	Sequence 66, Appl
14	29.8	36.8	653	10 US-09-923-779-53	Sequence 53, Appl
15	29.8	36.8	655	10 US-09-923-779-68	Sequence 68, Appl
16	29.8	36.8	656	10 US-09-923-779-69	Sequence 69, Appl
17	29.8	36.8	658	10 US-09-923-779-71	Sequence 71, Appl
18	29.8	36.8	659	10 US-09-923-779-52	Sequence 52, Appl
19	29.8	36.8	780	10 US-09-923-779-5	Sequence 5, Appl1

20	29.8	36.8	823	10 US-09-923-779-28	Sequence 28, Appl
21	29.8	36.8	88191	10 US-09-799-799-3	Sequence 3, Appl1
22	29.6	36.5	1619	9 US-09-764-868-400	Sequence 400, App
23	29.4	36.3	821	9 US-09-729-6588-7	Sequence 7, Appl1
24	29.4	36.3	1176	9 US-09-729-6588-14	Sequence 14, Appl
25	29.4	36.3	1574	9 US-09-729-6588-1	Sequence 1, Appl1
26	28.6	35.3	886	10 US-09-923-779-4	Sequence 4, Appl1
27	28.6	35.3	6158	10 US-09-919-497-6	Sequence 6, Appl1
28	28.6	35.3	6158	10 US-09-954-456-762	Sequence 762, App
29	28.4	35.1	1320	10 US-09-815-242-7615	Sequence 7615, Ap
30	27.6	34.1	572	10 US-09-923-779-56	Sequence 56, Appl
31	27.6	34.1	684973	10 US-09-263-959-1	Sequence 1, Appl1
32	27.2	33.6	274	10 US-09-923-876-4300	Sequence 4300, Ap
33	27	33.3	3913	9 US-10-026-188-7	Sequence 7, Appl1
34	27	33.3	155074	9 US-10-026-188-6	Sequence 6, Appl1
35	26.8	33.1	259	10 US-09-765-231A-49	Sequence 49, Appl
36	26.8	33.1	2235	10 US-09-962-436-308	Sequence 308, App
37	26.8	33.1	3226	10 US-09-954-456-725	Sequence 725, App
38	26.8	33.1	3690	12 US-10-044-090-448	Sequence 448, App
39	26.2	32.3	1133	9 US-09-924-340-57	Sequence 57, Appl
40	26.2	32.3	1133	9 US-09-992-600A-57	Sequence 57, Appl
41	26.2	32.3	2504	9 US-10-033-297-3	Sequence 3, Appl1
42	26.2	32.3	2504	9 US-10-081-806-3	Sequence 3, Appl1
43	26.2	32.3	2504	9 US-10-074-328-3	Sequence 3, Appl1
44	26.2	32.3	2511	10 US-09-777-430A-9	Sequence 9, Appl1
45	26.2	32.3	2511	10 US-09-777-430A-12	Sequence 12, Appl

ALIGNMENTS

RESULT 1
US-09-997-610-1
; Sequence 1, Application US/09997610
; Patent No. US20020156244A1
; GENERAL INFORMATION:
; APPLICANT: Fox, Brian
; APPLICANT: Holloway, James L.
; TITLE OF INVENTION: ADIPOCTE COMPLIMENT RELATED PROTEIN
; FILE REFERENCE: 00-96
; CURRENT APPLICATION NUMBER: US/09/997,610
; CURRENT FILING DATE: 2001-11-29
; PRIOR APPLICATION NUMBER: US 60/253,924
; PRIOR FILING DATE: 2000-11-29
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 1381
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (2)...(1381)
US-09-997-610-1
Query Match 100.0%; Score 81; DB 9; Length 1381;
Best Local Similarity 100.0%; Pred. No. 7.1e-18;
Matches 81; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GGAGCTCCAGACACCCAGGCCCCAGAGAGTGGGGCTCTGTCAGCAGGTTA 60
DB 56 GGAGCTCCAGACACCCAGGCCCCAGAGAGTGGGGCTCTGTCAGCAGGTTA 115
QY 61 CCACATATACAGAGAAATA 81
DB 116 CCACATATACAGAGAAATA 136
RESULT 2
US-09-997-610-5
; Sequence 5, Application US/09997610
; Patent No. US20020156244A1

GENERAL INFORMATION:
APPLICANT: Fox, Brian
APPLICANT: Holloway, James L.
TITLE OF INVENTION: ADIPOCYTE COMPLIMENT RELATED PROTEIN
FILE REFERENCE: 00-96
CURRENT APPLICATION NUMBER: US/09/997,610
CURRENT FILING DATE: 2001-11-29
PRIOR APPLICATION NUMBER: US 60/253,924
PRIOR FILING DATE: 2000-11-29
NUMBER OF SEQ ID NOS: 7
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 5
LENGTH: 1731
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (1)...(1731)
US-09-997-610-5

Query Match 100.0%; Score 81; DB 9; Length 1731;
Best Local Similarity 100.0%; Pred. No. 7.6e-18; Indels 0; Gaps 0;
Matches 81; Conservative 0; Mismatches 0;

QY 1 GGACCTCCAGACACCCAGGCCCCAGAGAGTGGGGCTCTGCTGACACAGGTTTA 60
DB 409 GGACCTCCAGACACCCAGGCCCCAGAGAGTGGGGCTCTGCTGACACAGGTTTA 468
QY 61 CCACAATATACAGAGAATA 81
DB 469 CCACAATATACAGAGAATA 489

RESULT 3
US-09-997-610-3
Sequence 3, Application US/09997610
Patent No. US200201562441
GENERAL INFORMATION:
APPLICANT: Fox, Brian
APPLICANT: Holloway, James L.
TITLE OF INVENTION: ADIPOCYTE COMPLIMENT RELATED PROTEIN
FILE REFERENCE: 00-96
CURRENT APPLICATION NUMBER: US/09/997,610
CURRENT FILING DATE: 2001-11-29
PRIOR APPLICATION NUMBER: US 60/253,924
PRIOR FILING DATE: 2000-11-29
NUMBER OF SEQ ID NOS: 7
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 3
LENGTH: 1377
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Degenerate polynucleotide encoding a polypeptide
OTHER INFORMATION: of SEQ ID NO:2
NAME/KEY: variation
LOCATION: (1)...(1377)
OTHER INFORMATION: Each n is independently A, T, G, or C.
NAME/KEY: misc_feature
LOCATION: (1)...(1377)
OTHER INFORMATION: n = A,T,C or G
US-09-997-610-3

Query Match 70.4%; Score 57; DB 9; Length 1377;
Best Local Similarity 64.2%; Pred. No. 6.7e-10; Indels 0; Gaps 0;
Matches 52; Conservative 9; Mismatches 20;

QY 1 GGACCTCCAGACACCCAGGCCCCAGAGAGTGGGGCTCTGCTGACACAGGTTTA 60
DB 55 GGACCTCCAGACACCCAGGCCCCAGAGAGTGGGGCTCTGCTGACACAGGTTTA 114

QY 61 CCACAATATACAGAGAATA 81
DB 115 CCACAATATACAGAGAATA 135

RESULT 4
US-09-997-610-7
Sequence 7, Application US/09997610
Patent No. US200201562441
GENERAL INFORMATION:
APPLICANT: Fox, Brian
APPLICANT: Holloway, James L.
TITLE OF INVENTION: ADIPOCYTE COMPLIMENT RELATED PROTEIN
FILE REFERENCE: 00-96
CURRENT APPLICATION NUMBER: US/09/997,610
CURRENT FILING DATE: 2001-11-29
PRIOR APPLICATION NUMBER: US 60/253,924
PRIOR FILING DATE: 2000-11-29
NUMBER OF SEQ ID NOS: 7
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 7
LENGTH: 1731
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Degenerate polynucleotide sequence of zacr13/znp1
OTHER INFORMATION: of SEQ ID NO:6
NAME/KEY: misc_feature
LOCATION: (1)...(1731)
OTHER INFORMATION: n = A,T,C or G
US-09-997-610-7

Query Match 70.4%; Score 57; DB 9; Length 1731;
Best Local Similarity 64.2%; Pred. No. 7.2e-10; Indels 0; Gaps 0;
Matches 52; Conservative 9; Mismatches 20;

QY 1 GGACCTCCAGACACCCAGGCCCCAGAGAGTGGGGCTCTGCTGACACAGGTTTA 60
DB 409 GGACCTCCAGACACCCAGGCCCCAGAGAGTGGGGCTCTGCTGACACAGGTTTA 468
QY 61 CCACAATATACAGAGAATA 81
DB 469 CCACAATATACAGAGAATA 489

RESULT 5
US-09-960-352-1786
Sequence 1786, Application US/09960352
Patent No. US20020137139A1
GENERAL INFORMATION:
APPLICANT: Warren, Wesley C.
APPLICANT: Tao, Nengsheng
APPLICANT: Byatt, John C.
APPLICANT: Mathialagan, Nagappan
TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
FILE REFERENCE: 16511.006/37-21(10298)C
CURRENT APPLICATION NUMBER: US/09/960,352
CURRENT FILING DATE: 2001-09-24
NUMBER OF SEQ ID NOS: 15112
SEQ ID NO 1786
LENGTH: 392
TYPE: DNA
ORGANISM: Bos taurus
OTHER INFORMATION: Clone ID: 08-LIB34-023-Q1-E1-B7
US-09-960-352-1786

Query Match 40.0%; Score 32.4; DB 10; Length 392;
Best Local Similarity 64.9%; Pred. No. 0.069; Indels 0; Gaps 0;
Matches 48; Conservative 0; Mismatches 26;

QY 1 GGACCTCCAGACACCCAGGCCCCAGAGAGTGGGGCTCTGCTGACACAGGTTTA 60

Db 155 GGCCACGACGAGACAGCCAGGCGCTCCAGGTATAGAGGGCGCTCCAGGTATAGAGGATTA 214
QY 61 CCACATATATACGAGG 74
Db 215 CCAGGTTTGCACAG 228

RESULT 6

US-09-960-352-9137
; Sequence 9137, Application US/09960352
; Patent No. US20020137139A1
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Tao, Nengbing
; APPLICANT: Byatt, John C.
; APPLICANT: Mathalagan, Nagappan
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
; FILE REFERENCE: 16511.006/37-21(10298)C
; CURRENT APPLICATION NUMBER: US/09/960,352
; CURRENT FILING DATE: 2001-09-24
; NUMBER OF SEQ ID NOS: 15112
; SEQ ID NO 9137
; LENGTH: 415
; TYPE: DNA
; ORGANISM: Bos taurus
; OTHER INFORMATION: Clone ID: 39-LIB34-018-Q1-E1-B12
US-09-960-352-9137

Query Match 40.0%; Score 32.4; DB 10; Length 415;
Best Local Similarity 64.9%; Pred. No. 0.07;
Matches 48; Conservative 0; Mismatches 26; Indels 0; Gaps 0;

QY 1 GGACTCCAGACACCCAGGCCCCAGAGAAGTGGGCGCTCTGTGCACGAGTTTA 60
Db 165 GGCCACGACGAGACCCAGGCGCTCCAGGTATAGAGGCGCTCCAGGTATAGAGGATTA 224
QY 61 CCACATATATACGAGG 74
Db 225 CCAGGTTTGCACAG 238

RESULT 7

US-09-923-779-49
; Sequence 49, Application US/09923779
; Patent No. US2002007621A1
; GENERAL INFORMATION:
; APPLICANT: Pyle, Ruth A.
; APPLICANT: Xu, Jiangchun
; APPLICANT: Kalos, Michael D.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.553
; CURRENT APPLICATION NUMBER: US/09/923,779
; CURRENT FILING DATE: 2001-08-06
; NUMBER OF SEQ ID NOS: 155
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 49
; LENGTH: 549
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 480, 498, 523, 539
; OTHER INFORMATION: n = A,T,C or G
US-09-923-779-49

Query Match 37.5%; Score 30.4; DB 10; Length 549;
Best Local Similarity 60.5%; Pred. No. 0.35;
Matches 49; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

QY 1 GGACCTCCAGCAGACCCAGGCCCCAGAGAAGTGGGCGCTCTGTGCACGAGTTTA 60

Db 406 GGAATCTCCAGATATACAGAGACCCCGCTGTGAACCTGGGCAAGCTGTCTTACAGGCCCT 465
QY 61 CCACATATATACGAGGAGATA 81
Db 466 CCAGAGCCTCTGCGCTATA 486

RESULT 8

US-09-880-107-2178
; Sequence 2178, Application US/09880107
; Patent No. US20020142981A1
; GENERAL INFORMATION:
; APPLICANT: Horne, Darci T.
; APPLICANT: Vockley, Joseph G.
; APPLICANT: Scherf, Uwe
; APPLICANT: Gene Logic, Inc.
; TITLE OF INVENTION: Gene Expression Profiles In Liver Cancer
; FILE REFERENCE: 44921-5028-WO
; CURRENT APPLICATION NUMBER: US/09/880,107
; CURRENT FILING DATE: 2001-06-14
; PRIOR APPLICATION NUMBER: US 60/211,379
; PRIOR FILING DATE: 2000-06-14
; PRIOR APPLICATION NUMBER: US 60/237,054
; PRIOR FILING DATE: 2000-10-02
; NUMBER OF SEQ ID NOS: 3950
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2178
; LENGTH: 3394
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Genbank Accession No. US20020142981A1 L22548
US-09-880-107-2178

Query Match 37.5%; Score 30.4; DB 10; Length 3394;
Best Local Similarity 71.4%; Pred. No. 0.6;
Matches 40; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

QY 1 GGACTCCAGCAGACCCAGGCCCCAGAGAAGTGGGCGCTCTGTGCACGAG 56
Db 670 GGCCCTCCAGAGACCCAGGCCCCAGAGAAGTGGGCGGCCCCCGGACACACAGG 725

RESULT 9

US-09-764-868-48
; Sequence 48, Application US/09764868
; Patent No. US2002016871A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT232
; CURRENT APPLICATION NUMBER: US/09/764,868
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - refer to PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1510
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 48
; LENGTH: 2686
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (2505)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (2569)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (2644)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (2645)

OTHER INFORMATION
US-09-923-779-47

```
; FEATURE:
; NAME/KEY: misc_feature
```


THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: February 19, 2003, 14:53:22 ; Search time 1157.56 Seconds

(without alignments)
2084.671 Million cell updates/sec

Title: US-09-997-610-2_COPY_1_149
Perfect score: 790
Sequence: 1 1V1PVLITAVLEHVEVAGP.....KEEISKOOSIQEVTWLLKA 149

Scoring table:
BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xih
-O=/cgn2_1/USPTO.spool/US09997610/runat_10022003_160823_23843/app-query.fasta.1.1635
-DB=EST -QFMT=fastap -SUFFIX=first -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=biosum62 -TRANS=human40.cdi -LIST=45
-DOCALLIGN=200 -THR.SCORE=pct -THR.MAX=100 -THR.MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pic -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US09997610_ECGN_1_1_181_6runat_10022003_160823_23843 -MCPU=6 -ICPU=3
-NO_XLPHY -NO_MMAP -LARGEQUERY -NEG.SCORES=0 -WAIT -LONGLOG -DEV.TIMEOUT=120
-WARN.TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -Fgapop=6 -Fgapext=7
-YGAPOP=10 -YGAPEXT=0.5 -DELop=6 -DELext=7

Database :
1: em_estba:*
2: em_esthum:*
3: em_estlin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hic:*
9: gb_est1:*
10: gb_est2:*
11: gb_hic:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_oth:*
26: em_gss_pro:*
27: em_gss_rtd:*

pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	311	39.4	499	14	BM967732
2	228.5	28.9	526	12	BE757276
3	226.5	28.7	504	13	BE757275
4	181	22.9	663	13	B068972
5	165.5	20.9	1092	17	CNS05408
6	158.5	20.1	512	13	B494317
7	158	20.0	587	10	AV593019
8	157.5	19.9	689	10	BB318780
9	157.5	19.9	2462	11	AK018742
10	157	19.9	662	14	BQ419957
11	155.5	19.7	695	9	AV243093
12	154.5	19.6	589	14	B0193813
13	154.5	19.6	921	9	AA673154
14	152.5	19.3	468	10	AW988554
15	152.5	19.3	552	10	BE625509
16	152.5	19.3	633	10	BB625284
17	152.5	19.3	707	12	BB872725
18	152.5	19.3	921	11	AK003138
19	151.5	19.2	460	9	A1332213
20	150.5	19.1	885	9	AA493090
21	150.5	19.1	508	9	A1323115
22	150	19.0	942	14	BQ900894
23	149	18.9	589	13	B1760895
24	148.5	18.8	856	14	BQ876848
25	147.5	18.7	1029	17	CBS022P5
26	146.5	18.5	607	12	BB691150
27	146	18.4	585	10	AW913280
28	145.5	18.4	580	12	BB691162
29	145.5	18.4	582	10	BE588278
30	145	18.4	430	10	AA453561
31	144.5	18.3	453	10	BE126709
32	144.5	18.3	639	13	B090235
33	142.5	18.0	693	9	AA060543
34	141.5	17.9	890	14	BQ737531
35	141	17.8	683	10	AW029321
36	140.5	17.8	529	12	BB074104
37	140	17.7	482	14	BQ260214
38	140	17.7	563	14	BQ449231
39	140	17.7	566	14	BQ285664
40	140	17.7	566	14	BQ258785
41	140	17.7	580	13	B1477496
42	140	17.7	587	14	BQ420027
43	140	17.7	597	14	BQ285786
44	140	17.7	598	14	BQ420183
45	140	17.7	626	14	BQ260132

ALIGNMENTS

RESULT 1
BM967732
LOCUS BM967732 499 bp mRNA linear EST 20-MAR-2002
DEFINITION LM24HM0134 Bos taurus LM-24-HW cDNA library Bos taurus cDNA clone
LM-24-HW-011-34 (5'), mRNA sequence.
ACCESSION BM967732
VERSION BM967732.1 GI:19561919
KEYWORDS EST.
SOURCE cow.
ORGANISM Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovidae; Bovinae; Bos.
REFERENCE
1 (bases 1 to 499)
Yoon,D.H., Jang,Y.S., Kim,T.H., Park,E.W., Lee,H.K., Chung,E.R.,

TITLE Sun, S.S. and Cheong, I.C.
JOURNAL Gene Expression Profiling of the Bovine skeletal muscle
COMMENT Unpublished (2002)
Contact: Dr. Du-Hak Yoon
National Livestock Research Institute, RDA
564 Omockchun-dong, Suwon, 441-350, Korea
Tel: 82 31 290 1593
Fax: 82 31 290 1792
Email: dhyoon@rda.go.kr
Insert length: 499 Std Error: 0.00
Seq primer: CAGGAACACGCTATGAC
POLYA-No.

FEATURES
Source Location/Qualifiers

1..499
/organism="Bos taurus"
/db_xref="taxon:9913"
/clone="LM-24-HW-011-34 (5')"
/clone_lib="Bos taurus LM-24-HW cDNA library"
/sex="six males mixed"
/tissue_type="longissimus dorsi"
/cell_type="myocyte"
/dev_stage="24 months old"
/lab_host="XLI-BlueMRF/strain"
/note="Organ: skeletal muscle; Vector: Uni-ZAPXR; Site:1;
EcoRI: Site:2: Xho I"

BASE COUNT 129 a 139 c 131 g 100 t
ORIGIN

Alignment Scores:

Pred. No.: 4,99e-23 Length: 499
Score: 311.00 Matches: 68
Percent Similarity: 63.49% Conservative: 12
Best Local Similarity: 53.97% Mismatches: 37
Query Match: 39.37% Indels: 9
DB: 14 Gaps: 3

US-09-997-610-2_COPY_1_149 (1-149) x BM67732 (1-499)

QY 1 l l e v a l l i l l e p r o v a l l e u l l e t h r a l a v a l l e g l u h i s ----- v a l g l u v a l 17
Db 120 C T G G T C A C A T C A T C A T G A A A T G C T G C T G A G A A A G A G G T A C C G A C C T 179
QY 18 A l a g l y p r o p r o a l a h i s p r o a r g p r o p r o g l u g l u a l g l y p r o p r o ----- g l y 34
Db 180 C C G G G C C A C A G A C A C C A G G C C T C A G G T A A A G A G G C C T C A G T A A G A G A 239
QY 35 A l a p r o g l y l e u p r o g l i n t y r t h r g l y l u l l e s e r g l u m e t h r l y s c y s p r o c y s p r o 54
Db 240 A T A C C A G G T T T G C C A G G T C C G C G G A A C T C A G A C A A G T A A A T G C C C A T G C C A C 299
QY 55 A s p l i e g l u a r g s e r a l a p h e r t h r v a l l y l s l e u s e r g l y l y l s l e u p r o l e u p r o p h e l y s 74
Db 300 ----- A G A C A G C T C C C C T T C A C T G T A A G C T C A G T G G C C A G T T C C C C T T C A A A G 353
QY 75 P r o i l e - l l e p h e r t h r g l y v a l l e u t y r a s n a l a g l i n a r g a s p l e u l y s g l u a l a m e t g l 94
Db 334 C C T G T G C C C T T T C A G A G G T C T G T A C A A T G C C A G A A A G A C T T C A C A G A G A C A C T G G 413
QY 94 y v a l p h e a l a c y s a r g v a l p r o g l y a s n t y r t y r s e r s e r p h e a s p v a l g l u l e u h i s h i 114
Db 414 G G T C T T C A C A T G A C A G G T G C C A G A A T T A C A T T T C C T T T T A T G T G A T C T C A T C A 473
QY 114 s c y l s y l v a l a s n i l e 119
Db 474 C T G C A A G G T G A C T G T T 489

RESULT 2
BE757276 526 bp mRNA linear EST 25-APR-2001
LOCUS BE757276
DEFINITION 211669 MARC 2BOV Bos taurus cDNA 5', mRNA sequence.
ACCESSION BE757276
VERSION BE757276.1 GI:10171268
KEYWORDS EST.

SOURCE
ORGANISM cow.
Bos taurus
Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovidae; Bovinae; Bos.
REFERENCE
AUTHORS 1 (bases 1 to 526)
Smith, T.P.L., Grosse, W.M., Freking, B.A., Roberts, A.J., Stone, R.T.,
Casas, E., Wray, J.E., White, J., Cho, J., Fahrenkrug, S.C., Bennett,
G.L., Heaton, M.P., Laegreid, W.W., Rohrer, G.A., Chitko-Mckown, C.G.,
Pietra, G., Holt, I., Karamycheva, S., Llang, F., Quackenbush, J. and
Keefe, J.W.
Sequence evaluation of four pooled-tissue normalized bovine cDNA
libraries and construction of a gene index for cattle
Genome Res. 11 (4), 626-630 (2001)

TITLE
JOURNAL
MEDLINE
COMMENT
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390
Email: smtlh@email.marc.usda.gov
Single pass sequencing. Bases called and alt-trimmed with Phred
v0.980904.e. Vector identified by cross-match with the -m1nscore 18
and -m1match 12 options.
PCR primers
FORWARD: AGGAACACGCTATGACCAT
BACKWARD: GTTTCACGTCACGACG
Plate: 63 row: 6 column: 14
Seq primer: ATTGAGTACACCTATTA.

FEATURES
Source Location/Qualifiers

1..526
/organism="Bos taurus"
/db_xref="taxon:9913"
/clone_lib="MARC 2BOV"
/tissue_type="pooled"
/lab_host="DH10B"
/note="Vector: pCMV SPORT6; Site:1: NotI; Site:2: SalI;
Library made from pooled tissue from testis, thymus,
semilandinosus muscle, longissimus muscle, pancreas,
adrenal, and endometrium."

Alignment Scores:

Pred. No.: 2,93e-14 Length: 526
Score: 228.50 Matches: 52
Percent Similarity: 55.26% Conservative: 11
Best Local Similarity: 45.61% Mismatches: 40
Query Match: 28.92% Indels: 11
DB: 12 Gaps: 3

US-09-997-610-2_COPY_1_149 (1-149) x BE757276 (1-526)

QY 19 G l l y p r o p r o ----- A l a h i s p r o a r g p r o p r o g l u g l u a l g l y p r o g l y a l a 35
Db 181 G G A C C T C C G G A C T A G T G T C C A G G C C A C C G T T A C A G A C C A T A G A T G 240
QY 36 P r o g l y l e u ----- P r o g l i n t y r t h r g l y l u l l e s e r g l u m e t h r l y s c y s p r o 52
Db 241 C C A G A C T A A C A G G A G A C T G G C T C C T G A C A G T G T T A G A A G ----- 285
QY 53 C y s p r o a s p l i e g l u a r g s e r a l a p h e r t h r v a l l y l s l e u s e r g l y l y l s l e u p r o l e u p r o 72
Db 286 T G C C A C C C C T G C T C A G T C C T T T C C G T T C C A A G C G A G T G G C C T T T C C A G A C C C 345
QY 73 P h e l y s p r o i l e l l e p h e r t h r g l y v a l l e u t y r a s n a l a g l i n a r g a s p l e u l y s g l u a l a 92
Db 346 T C C A C A C C A T T G T C T C A G A A G T T C T G A C A C A T C A G T G G C C A C T T G A C A C C C G C C 405
QY 93 M e t g l y v a l p h e a l a c y s a r g v a l p r o g l y a s n t y r t y r s e r s e r p h e a s p v a l g l u l e u 112
Db 406 A C T G T G T T C A G C T G C A G C C T C C T G T G T A C C A C T T T G C C T T T G A C A T T G A C T T G 465

OY 113 H1SH1SCysLysValAsnIleTrrPleuMetArgLysGlnIle 126
 Db 466 TTTCAGAGTGTCTCAAGGTGGCTTAATGCGAATGCGATC 507
 RESULT 3
 BE757275 504 bp mRNA linear EST 25-APR-2001
 LOCUS 211668 MARC 2BOV Bos taurus cDNA 5', mRNA sequence.
 DEFINITION BE757275
 ACCESSION BE757275
 VERSION BE757275.1 GI:10171267
 KEYWORDS EST.
 SOURCE cow.
 ORGANISM Bos taurus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 Bovidae; Bovinae; Bos.
 REFERENCE 1 (bases 1 to 504)
 AUTHORS Smith,T.P.L., Grosse,W.M., Freking,B.A., Roberts,A.J., Stone,R.T.,
 Casas,E., Wray,J.E., White,J., Fahrenkrug,S.C., Bennett,
 G.L., Heaton,M.P., Laegreid,W.W., Rohrer,G.A., Chitko-McKown,C.G.,
 Pettes,G., Holt,I., Karanycheva,S., Liang,F., Quackenbush,J. and
 Keefe,J.W.
 TITLE Sequence evaluation of four pooled-tissue normalized bovine cDNA
 libraries and construction of a gene index for cattle
 JOURNAL Genome Res. 11 (4), 626-630 (2001)
 MEDLINE 21180013
 COMMENT Contact: Smith TPL
 USDA, ARS, US Meat Animal Research Center
 PO Box 166, Clay Center, NE 68933-0166, USA
 Tel: 402 762 4366
 Fax: 402 762 4390
 Email: smithemall.marc.usda.gov
 Single pass sequencing. Bases called and alt_trimmed with phred
 v0.980904.e. Vector identified by cross_match with the minscore 18
 and -mismatch 12 options.
 PCR PRIMERS
 FORWARD: AGGAACAGCTATGACCAT
 BACKWARD: GTTTCAGTCCAGCGACG
 Plate: 63 row: G column: 13
 Seq primer: ATTTAGGTGACATTAAG.
 FEATURES
 source Location/Qualifiers
 1..504
 /organism="Bos taurus"
 /db_xref="taxon:9913"
 /clone_lib="MARC 2BOV"
 /tissue_type="pooled"
 /lab_host="DH10B"
 /note="Vector: PCMV SPORT6; Site.1: NotI; Site.2: SalI;
 library made from pooled tissue from testis, thymus,
 semitendinosus muscle, longissimus muscle, pancreas,
 adrenal, and endometrium."
 BASE COUNT 99 a 131 c 146 g 128 t
 ORIGIN
 Alignment Scores:
 Pred. No.: 4.49e-14 Length: 504
 Score: 226.50 Matches: 51
 Percent Similarity: 55.86% Conservative: 11
 Best Local Similarity: 45.95% Mismatches: 38
 Query Match: 28.67% Indels: 11
 DB: 12 Gaps: 3
 US-09-997-610-2_COPY_1_149 (1-149) x BE757275 (1-504)
 OY 19 GlyProPro-----AlaHisProArgProProGluGluValGlyProProGlyAla 35
 Db 181 GGACCTCGCGGACCTGATGCTTCACAGGCCACCTGTTATACAGACCATAGTATG 240
 OY 36 ProGlyLeu-----ProGlnTyrThrGlyGluIleSerGlnMetThrLysCysPro 52
 Db 241 CCAGAGCAACAGAGGAGACCTGGCTTCCTGAGCTGTTGAGAG----- 285
 OY 53 CysProAspIleGluArgSerAlaPheThrValLysLeuSerGlyLysLeuProLeuPro 72

Db 286 TCCCCACCCCTGCCTCAGCTGCTGCTTCCGTCACAGCTGAGGGCCTTCCACAGACCC 345
 OY 73 PheLysProIleIlePheThrGlyValLeuTyrAsnAlaGlnArgAspLeuLysGlnAla 92
 Db 346 TCCACAGCCCATTTGCTTCACAGAAAGTTCGTCAACACCACTGACCATTCGACACCCGCC 405
 OY 93 MetGlyValPheAlaCysArgValProGlyAsnTyrTyrSerSerPheAspValGluLeu 112
 Db 406 ACTGCTGTTCAGCTGACGACGCTCCCTGCTGTACACACTTGGCTTGCATTCAGTTG 465
 OY 113 H1SH1SCysLysValAsnIleTrrPleuMetArg 123
 Db 466 TTTCAGAGTGTCTCAAGGTGGCTTAATGCGG 498
 RESULT 4
 BJO68972 663 bp mRNA linear EST 11-DEC-2001
 LOCUS BJO68972 NIBB Mochil normalized Xenopus tailbud library Xenopus
 DEFINITION laevis cDNA clone XJ051f08 5', mRNA sequence.
 ACCESSION BJO68972
 VERSION BJO68972.1 GI:17497722
 KEYWORDS EST.
 SOURCE African clawed frog.
 ORGANISM Xenopus laevis
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;
 Xenopodinae; Xenopus.
 REFERENCE 1 (bases 1 to 663)
 AUTHORS Kitayama,A., Terasaka,C., Mochi,M., Ueno,N., Shin-I,T. and Kohara
 Y.
 TITLE Expressed genes in X. laevis embryo
 JOURNAL Unpublished (2001)
 COMMENT Contact: Tadasu Shin-I
 Center For Genetic Resource Information
 National Institute of Genetics
 1111 Yata, Mishima, Shizuoka 411-8540, Japan
 Tel: 81-559-81-6856
 Fax: 81-559-81-6855
 Email: tshini@genes.nig.ac.jp.
 FEATURES
 source Location/Qualifiers
 1..663
 /organism="Xenopus laevis"
 /db_xref="taxon:8355"
 /clone_lib="XJ051f08"
 /clone_lib="NIBB Mochil normalized Xenopus tailbud
 library"
 /tissue_type="whole embryo"
 /dev_stage="stage 25"
 /note="Vector: pBSRN3; Site.1: NotI; Site.2: EcoRI; cDNAs
 were oligo-dT primed and directionally cloned. Staging
 according to Faber, Nieuwkoop and Faber. Library is subtracted
 and was constructed by N. Garrett and A.M. Zorn,
 (Wellcome/CRC Institute)."
 BASE COUNT 179 a 160 c 164 g 160 t
 ORIGIN
 Alignment Scores:
 Pred. No.: 4.37e-09 Length: 663
 Score: 181.00 Matches: 46
 Percent Similarity: 42.55% Conservative: 14
 Best Local Similarity: 32.62% Mismatches: 47
 Query Match: 22.91% Indels: 34
 DB: 13 Gaps: 6
 US-09-997-610-2_COPY_1_149 (1-149) x BJO68972 (1-663)
 OY 18 AlaGlyProProAlaHisProArgProProGluGluValGlyProProGlyAlaProGly 37
 Db 106 TCAGGGCCACCTGATGCTGATGACAGACCTCA-----GGACACACAGAGCAACCTGGC 156
 OY 38 LeuPro-----GlnTyrThrGlyGluLysSer 46

```

Db      157  CGTCCTGCAGAAATCGAGACCTACGCCCCAGGATGAGCAGATTTGTCACAAATATGCGA 216
Oy      47  GUnethrlyscysProasp-----55
Db      217  GCCATTGATGAGTAAACACCTCATGAGTATGACGAGGAAAAAAGAAATGAGGA 276
Oy      56  -----1legUARserAlapherthValLysLeuSerGlyLysLeuProleuProPhe 73
Db      277  AACATTCACGAATGCCAGCTCAGCTGCAAGCTGACGCCCTTCCTTATGTCGT 336
Oy      74  LysProlellepherthGlyValLeuTyAsnAlaGlnArgAspLeuLysGluAlaMet 93
Db      337  GAACCTATTAAGTTGAGAAACCTGTATACACGACGAAATGGCTACAAATCCACTTACT 396
Oy      94  G1ValPheAlaCysArgValProGlyAsnTyTyrSerSerPheAspValGluLeuHis 113
Db      397  GGGGTTCACGTGTGAATTCAGGAATTTATTTACTTTTCTTACCATGTC-----447
Oy      114  HisCysLys---ValAsnLleTrrpleuMetArgLysGlnLleuAlaAsnLysGlu 132
Db      448  CATGTAAAGTGCCAAATGTGTGAT-----GCGTTGTTCAAGAAATATGAGCCA 498
Oy      133  Ile 133
Db      499  GTA 501

```

```

RESULT 5
CNS0540B 1092 bp DNA linear GSS 26-JUL-2000
LOCUS Tetradon nigroviridis genome survey sequence 73 end of clone
DEFINITION 011D22 of library A from Tetradon nigroviridis, genomic survey
sequence.
ACCESSION AL320996
VERSION AL320996.1 GI:9553880
KEYWORDS GSS: genome survey sequence.
SOURCE Tetradon nigroviridis.
ORGANISM Tetradon nigroviridis.
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
Tetraodontidae; Tetradon.

```

```

REFERENCE
AUTHORS 1 (bases 1 to 1092)
Roest Crollius,H., Jalllon,O., Dasliya,C., Ozouf-Costaz,C., Fitzames,C.,
Bernot,A., Fitzames,C., Winkler,P., Brotlier,P., Quetier,F.,
Saurin,W. and Weissenbach,J.
ESTimate of human gene number provided by genome-wide analysis
using Tetradon nigroviridis DNA sequence
JOURNAL Nat. Genet. 25 (2), 235-238 (2000)
MEDLINE 20296633
PUBMED 10835645
TITLE 2 (bases 1 to 1092)
Crollius,H.R., Jalllon,O., Dasliya,C., Ozouf-Costaz,C., Fitzames,C.,
Fischer,C., Bouneau,L., Billault,A., Quetier,F., Saurin,W.,
Bernot,A. and Weissenbach,J.
Characterization and repeat analysis of the compact genome of the
freshwater pufferfish Tetradon nigroviridis
JOURNAL Genome Res. 10 (7), 939-949 (2000)
MEDLINE 20359837
PUBMED 10899143
TITLE 3 (bases 1 to 1092)
Genoscope.
REFERENCE Direct Submission
AUTHORS Submitted (12-APR-2000)
JOURNAL This sequence is a single read and was generated as part of a large
COMMENT scale clone-end sequencing project of the Tetradon nigroviridis
genome. For more information, please take a look at
http://www.genoscope.cns.fr/Tetradon.
location/Qualifiers
FEATURES
source 1..1092
/organism="Tetradon nigroviridis"
/db_xref="taxon:99883"
/clone="011D22"
/clone_1db="A"

```

```

BASE COUNT 224 a 301 c 308 g 245 t 14 others
ORIGIN

```

```

Alignment Scores:
Pred. No.: 3.9e-07
Score: 165.50
Percent Similarity: 45.80%
Best Local Similarity: 32.06%
Query Match: 20.95%
DB: 17

```

```

US-09-997-610-2_COPY_1_149 (1-149) x CNS0540B (1-1092)

```

```

Oy      19  G1YProProlAlaHisProArgProProGluGluValGlyProProGlyAla-----35
Db      528  GGGGCCCTCGGCCACGCTGTGAGCGCTGTGAGCGTGTGAGCGAAGTGTGTGTGAGGAAATGCGTGTGAG 587
Oy      36  ProGlyLeuProGlnTyrThrGlyGluLe-----Ser 46
Db      588  CCGGACCGCGCTGCTCTCTCTCTGCGAAGTGTGTGTGTGAGGAAATGCGTGTGAG 647
Oy      47  GUnethrlyscysProaspProaspPileGluArgSerAlapherthValLysLeuSer 66
Db      648  GTTATGGTCAAATGCC-----ATGTCCTGCTTCTACCTGCTCTCTGAC 692
Oy      67  GlyLysLeuProleuProPheLysProlellepherthGlyValLeuTyAsnAlaGln 86
Db      693  ACACCTACCCAGCTGCTGTGACGCTGTGACGCTGTGACGCTGTGACGCTGTGACGCTGTGAG 752
Oy      87  ArgAspLeuLysGluAlaMetGlyValPheAlaCysArgValProGlyAsnTyTyrSer 106
Db      753  AATCAGTATTAACCCGAGCTGTGACGCTGTGACGCTGTGACGCTGTGACGCTGTGACGCTGTG 812
Oy      107  SerPheAspValGluLeuHis-----HisCysLysValAsnLleTrrpleuMetArgLys 124
Db      813  TCTTACAGATCATGTCATGTCATGTCATGTCATGTCATGTCATGTCATGTCATGTCATGTCAT 872
Oy      125  GlnLleuAlaAsnLysGluGluLysSerLys 135
Db      873  CCTATTATGTTCACTTATGACGATACACACAG 905

```

```

RESULT 6
BU494317 512 bp mRNA linear EST 08-AUG-2002
LOCUS BU494317 MF01FSA cDNA Oryzias latipes cDNA clone MF01FSA024E16 5'
DEFINITION mRNA sequence.
ACCESSION BU494317
VERSION BU494317.1 GI:22146243
KEYWORDS EST.
SOURCE Japanese medaka.
ORGANISM Oryzias latipes.
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthomorpha; Acanthopterygii; Percomorpha; Altheriomorpha;
Belontiiformes; Adrianchthyidae; Oryziinae; Oryzias.

```

```

REFERENCE 1 (bases 1 to 512)
AUTHORS Kohara,Y., Shin-I,T., Kimura,T., Marita,T., Jindo,T. and Takeda,H.
TITLE Medaka EST Project in Takeda's lab
JOURNAL Unpublished (2001)
COMMENT Contact: Tadasu Shin-1
Center for Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tshin@genes.nig.ac.jp.
location/Qualifiers
FEATURES
source 1..512
/organism="Oryzias latipes"
/strain="ig-tr"
/db_xref="taxon:8090"
/clone="MF01FSA024E16"

```

[illegible]

Single pass sequencing.
This clone was obtained from a polyA-deleted cDNA library.

```

FEATURES
    source
        .1..587
            /organism="Bos taurus"
            /db_xref="taxon:9913"
            /clone_1ib="Bos taurus cartilage fetus"
            /tissue_type="cartilage"
            /dev_stage="fetus"
            /lab_host="DH10B"
            /note="Vector: pLL1; Site_1: SalI; Site_2: NotI; Poly A
was deleted from a NotI site"
BASE COUNT      114 a      161 c      170 g      138 t      4 others
ORIGIN
Alignment Scores:
Pred. No.:       1e-06          Length:         587
Score:           158.00         Matches:         40
Percent Similarity: 45.61%     Conservative:   12
Best Local Similarity: 35.09%   Mismatch:       44
Query Match:      20.00%       Indels:         18
DB:               10          Gaps:             4
US-09-997-610-2_COPY_L_149 (1-149) x AV593019 (1-587)
QY      18 AAGAGTProProlAHISProArGrProFrogUGluGValAGlyProProGlyVALAProGly 37
        |||||  ||  |||||  ||  |||||  ||  |||||  ||  |||||  ||  |||||  ||
Db      347 GCTGGGGAAGCCTGTGCCTCCACAGGCCGCCACAGGCCGCCACAGGCCCTCCAGCAACACC 288
QY      38 LeuPProgin-----tyrThrGlyGUILeserGIunet 48
        |||||||:::
Db      287 CTGCCCGAGGACTTTGTAAAGCACAGCCCAAAGGCCGTTTGNNAGTCACATCAGGAGTA 228
QY      49 ThrLyScysProGcysProAsprIdeLuarGSerAlaPheThyValLeuLeuSerGlylys 68
        |||||:::
Db      227 ACAGGATGCTGTG-----TCGTCTTACCTGTTATCTCTCCAAAGCT 183
QY      69 LeuProlauproPhelaysProIlelePheThrcylValLeuTyFAANAglinarasp 88
        |||||  ||  |||||  ||  |||||  ||  |||||  ||  |||||  ||  |||||  ||
Db      182 TACCACGCTATTATAGTACCTCATTCCTCATTCGATTGATTAATATACAGCAACAGCAT 123
QY      89 LeuLySGUalAmetGlyValPheAlacysarGvalAProGlyAsnTyrtTyserSerPhe 108
        |||||:::
Db      122 TATGACCCACAGAACCTGGATCTTCACCTGTAGAATTCCAGGGCATATTACTTCTTAC 63
QY      109 AspvAlaGUleuHSHISCyslys---ValasnIletrpleu 121
        ::::|  ||  |||||  ||  |||||  ||  |||||  ||  |||||  ||  |||||  ||
Db      62 CACATA-----CACGTGAAGAGGACCCACTGCTTGGGTA 30
RESULT 8      BB318780          689 bp      mRNA      linear      EST 24-OCT-2001
LOCUS      BB318780
DEFINITION  BB318780 RIKEN full-length enriched, adult male corpora
quadrifemina Mus musculus cDNA clone B2J3077C02 3, mRNA sequence.
ACCESSION  BB318780
VERSION    BB318780.2 GI:16403219
KEYWORDS   EST.
SOURCE     house mouse.
ORGANISM   Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Musineae; Mus.
1 (bases 1 to 689)
REFERENCE   1 (bases 1 to 689)
AUTHORS   Arikawa,T., Carolinci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A.,
Hiramoto,K., Horii,F., Ishii,Y., Ito,M., Kawachi,J., Komono,H., Kouda,
'M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M.,
Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasakih,
'D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H.,
Tezumi,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toya,T.,
Muramatsu,M. and Hayashizaki,Y.
RIKEN Mouse ESTs (Arikawa,T., et al. 2001)
Unpublished (2001)
On Jul11, 2000 this sequence version replaced gi:9025815.

```

Contact: Yoshinide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic
Sciences Center(GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsr.riken.go.jp,
URL: http://genome-gsc.riken.go.jp/

Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh
M., Kono, H., Okazaki, Y., Muramatsu, M., and Hayashizaki, Y.
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new
genes. Genome Res. 10 (10), 1617-1630 (2000)

Wagi, K., Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E.,
Watanabe, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura
S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A., and
Hayashizaki, Y.

RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer. Genome Res.
10 (11), 1757-1771 (2000)

Kono, H., Fukunishi, Y., Shibata, K., Itoh, M., Carninci, P., Sugahara
Y., and Hayashizaki, Y.

Computer-based methods for the mouse full-length cDNA
encyclopedia: real-time sequence clustering for construction of a
nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)

Kondo, S., Shinagawa, A., Saito, T., Kiyosawa, H., Yamanka, I., Alizawa
K., Fukuda, S., Hara, A., Itoh, M., Kawai, D., Shibata, K., and
Hayashizaki, Y.

Computational Analysis of Full-length Mouse cDNAs Compared with
Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)

Please visit our web site (<http://genome-gsc.riken.go.jp/>) for
further details.

cDNA library was prepared and sequenced in Mouse Genome
Encyclopedia Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in
Division of Experimental Animal Research in Riken contributed to
prepare mouse tissues

FEATURES

source

Location/Qualifiers
1. 689
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="B23037C02"
/clone_lib="RIKEN full-length enriched, adult male corpora
quadrigenina"
/sex="male"
/tissue_type="corpora quadrigenina"
/dev_stage="adult"
/lab_host="DH10B"
/note="Site_1: Sall; Site_2: BamHI; cDNA library was
prepared and sequenced in Mouse Genome Encyclopedia
Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in
RIKEN. Division of Experimental Animal Research in Riken
contributed to prepare mouse tissues. 1st strand cDNA was
primed with a primer [5'
GAGAGAGAGAGATCCAGAGCTCTTTTCTTTTCTTTTCTTTVN 3'], cDNA was
prepared by using trehalose thermo-activated reverse
transcriptase and subsequently enriched for full-length by
cap-trapper. cDNA went through one round of normalization
to Rot = 20.0 and subtraction to Rot = 459.0. Second
strand cDNA was prepared with the primer adapter of
sequence [5' GAGAGAGAGATCTCGATTAATTAATTAATCCCCCCCCCC
3']. cDNA was cleaved with XhoI and BamHI. Vector: a
modified pBluescript KS(+) after bulk excision from Lambda
PLC I."

BASE COUNT 169 a 189 c 182 g 147 t 2 others

ORIGIN

Alignment Scores:
Score: 1.42e-06 Length: 689
157.50 Matches: 40
Percent Similarity: 43.20% Conservative: 14

Best Local Similarity: 32.00% Mismatches: 45
Query Match: 19.94% Indels: 25
Db: 10 Gaps: 4

us-09-997-610-2_copy_1_149 (1-149) x BB318780 (1-689)

Qy 19 GYpProbaLhnsProArgrProoSLuGluValLgLYpProProGlyAlAProGlyLeu 38

Db 89 GGCAGCGCTGGCCCTTCCTGAGACCCCGAGGCTCCTCGAGACCCCGAGGCTG 148

Qy 39 -----ProGlnTyrThrGlyGlu----- 44

Db 149 ATGCTTACACATCACCCAGGAGAGAGTATCGCAGATATGAGACTAGATGATGG 208

Qy 45 -----IleSerGluMetThrlyCysProCysProoSLpIleGlu 57

Db 209 GTGAACCTCCGCATGCTATGCGGCAAAAGGCAACACGAGGCGCAGCTATGAG 268

Qy 58 ArgSerLaherTherVallyLysLeuSerGlyLysLeuProLeuProPhelyProIleIle 77

Db 269 ATGCTGCGGTTCCTCCGAGCTGACTGATACCTTTCCACGCGTGGGCGCCAGTGAAG 328

Qy 78 PheThrGlyValleuTyrAsnAlaGlnArqAspLeuLysGluAlaMetGlyValPheAla 97

Db 329 TTTGACAGCTGCTCTTCAACAGCGACAGACTACATCCGACAGAGCGATCTTACC 388

Qy 98 CysArgValProGlyAsnTyrSerSerPheAspValGluLeuHSHsCysLys--- 116

Db 389 TGTGAAGTCCCGGGGTCTACTACTTTCCTTATCAGTGT-----CACTGCAGAGGA 439

Qy 117 ValAsnIleThrLeu 121

Db 440 GGCACGATATGGGT 454

RESULT 9

AK018742

LOCUS

DEFINITION

AK018742 2462 bp mRNA linear HTC 19-JAN-2002

Mus musculus adult male kidney cDNA, RIKEN full-length enriched

library, clone:06100111.5; procollagen, type VIII, alpha 1, full

insert sequence.

AK018742

AK018742.1 GI:12858612

HTC; CAP trapper.

Mus musculus (strain:C57BL/6J) adult male kidney cDNA to mRNA,

clone:06100111.5.

ORGANISM

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE

1 Carninci, P., and Hayashizaki, Y.

High-efficiency full-length cDNA cloning

meth. Enzymol. 303, 19-44 (1999)

JOURNAL

MEDLINE

PUBMED

10343636

2

Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,

Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M., and Hayashizaki, Y.

Normalization and subtraction of cap-trapper-selected cDNAs to

prepare full-length cDNA libraries for rapid discovery of new genes

Genome Res. 10 (10), 1617-1630 (2000)

JOURNAL

MEDLINE

PUBMED

2049374

11042159

3

Shibata, K., Itoh, M., Alizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,

Kono, H., Akiyama, J., Nishi, K., Kitsumaki, T., Tashiro, H., Itoh, M.,

Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishino, T., Harada, A.,

Yamamoto, R., Watanabe, H., Sakaguchi, S., Ikegami, T., Keshiwa, K.,

Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watanabe, K.,

Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J.,

Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A., and Hayashizaki, Y.

RIKEN integrated sequence analysis (RISA) system--384-format

sequencing pipeline with 384 multicapillary sequencer

ALIGNMENT SCORES:	2,36e-06	Length:	589
Pred. No.:	154.50	Matches:	42
Score:	41.04%	Conservative:	13
Percent Similarity:	31.34%	Mismatches:	40
Best Local Similarity:	19.56%	Indels:	39
Query Match:	14	Gaps:	6
DB:			
US-09-997-610-2_COPY_1_149 (1-149) x BQ193813 (1-589)			
QY 14 H1sValGIuValaLagIyProProaIaH1sProaRgProProGIuGIuValaGIyProPro 33			
DB 587 CACGAGGCCCTCTCTGGCCACCAGGCCCTCCAGGCCCA-----G6CCCCCA 537			
QY 34 GlyAla---ProGlyLeuProGIuInrYrhrGluIeserGIuIethrIlyscsPro 52			
DB 536 GCTGTGATGCCCTACACCATCACCACCCAGGAGAGTAT----- 498			
QY 53 CysProAspIle----- 56			
DB 497 CTACACAGATATAGGACTAGGAATTCACATCCAGACCTCCACATGCCATYGGCGGCAAA 438			
QY 57 -----GluAsrAlaPheThrValIyIleuSerGIyIys 68			
DB 437 AACGGCAACAGAGGAGGCCCTATGAGATGCCGGGTTTACTGCCAGCTGACTACT 378			
QY 69 LeuProIeuProPheIyProIleIlePheThrGIyValIeYrAsnAlaGIuATGAsp 88			
DB 377 TTCACACGGGTGGGGCCCGACGTATGACAGGCTTATCAATGAGACAGCAAC 318			
QY 89 LeuIySGluAlaMetGIyValPheAlaCysATGValProGIyAsnTYrTYrSerSerpe 108			
DB 317 TACAAACCGCAGACGGAGCTTACCTCGGAAGATCCGGGTGTCTACTACTTGTCTTAC 258			
QY 109 AspValGIuLeuH1sCysIyS---ValAsnIleTyrIeu 121			
DB 257 CATCTT-----CACTGCACAGGAGGACACAGCTGTGGGT 225			
RESULT 13		921 bp	mRNA linear EST 26-NOV-1997
LOCUS	AA673154		
DEFINITION	ve40g07.r1 Soares_mammari_gland_MNMKG Mus musculus cDNA clone IMAGE:820668 5', similar to gp:037222 Mus musculus 30kDa adipocyte complement-related protein Acrp30 (MOUSE);, mRNA sequence.		
ACCESSION	AA673154		
VERSION	AA673154.1	GI:2646436	
KEYWORDS	EST.		
SOURCE	house mouse.		
ORGANISM	Mus musculus		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
AUTHORS	1 (bases 1 to 921)		
TITLE	Marré,M., Hillier,L., Allen,M., Bowles,M., Dierrich,N., Dubuque,T., Geisel,S., Kucaba,T., Lacy,M., Le,M., Matlin,J., Morris,M., Scheilenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B., Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and Waterston,R.		
JOURNAL	The WashU-HMNI Mouse EST Project		
COMMENT	Unpublished (1996) Contact: Marra M/Mouse EST Project WashU-HMNI Mouse EST Project Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 Tel: 314 266 1800 Fax: 314 266 1810 Email: mouseest@wustl.edu This clone is available royalty-free through LBNL; contact the IMAGE Consortium (info@image.lbnl.gov) for further information.		

151 t

MGI:488948
 Seq primer: -28m3 rev2 ET from Amersham
 High quality sequence stop: 499.

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus.p2n model

Run on: February 19, 2003, 14:57:18 ; Search time 627.004 Seconds
(without alignments)
2088.706 Million cell updates/sec

Title: US-09-997-610-2_COPY_1_45
Perfect score: 240
Sequence: 1 IVVPLVITAVIEHVEVAGP.....PPEVGPAGAGLPQYTGEEI 45

Scoring table:
BLOSUM62
Xgapop 10.0 , Ygapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2054640 seqs, 14551402878 residues
Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame_plus.p2n model -DEV=xlh
-Q=/cgn2_1/USPIO.spool/US09997610/runcat_10022003_160823_23819/app_query.fasta.1.1635
-DB=GenDbml -QFMT=fastap -SUFFIX=rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=Dlosum62 -TRANS=human40.cdi -LIST=45
-DOCLALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=plc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09997610.cgn2_1_13965.crunat_10022003_160823_23819 -NCPU=6 -ICPU=3
-NO_XLPXY -NO_MMAB -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV=TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -Fgapop=6 -Fgapext=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : GenEmbl:*
1: gb_ba:*
2: gb_hcg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*

29: em_vl:*
30: em_hcg_hum:*
31: em_hcg_inv:*
32: em_hcg_other:*
33: em_hcg_mus:*
34: em_hcg_pin:*
35: em_hcg_rtd:*
36: em_hcg_mam:*
37: em_hcg_vrt:*
38: em_sy:*
39: em_hcgo_hum:*
40: em_hcgo_mus:*
41: em_hcgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	197	82.1	145880	9 HS302D9	282198 Human DNA s
2	89	37.1	66669	3 AF017777	AF017777 Drosophill
3	89	37.1	87089	2 AC013980	AC013980 Drosophill
4	89	37.1	171548	3 AC092494	AC092494 Drosophill
5	89	37.1	173508	3 AC011251	AC011251 Drosophill
6	89	37.1	292911	3 AE003568	AE003568 Drosophill
7	87.5	36.5	68661	3 AC024805	AC024805 Caenorhab
8	87.5	36.5	278007	2 AC006799	AC006799 Caenorhab
9	86	35.8	4428	6 AX146422	AX146422 Sequence
10	86	35.8	4428	6 AX146424	AX146424 Sequence
11	85.5	35.6	5358	5 AT052763	AT052763 Xenopus 1
12	85	35.4	256417	2 AC079422	AC079422 Mus muscu
13	84	35.0	4428	6 AX146430	AX146430 Sequence
14	84	34.6	186	6 AR014097	AR014097 Sequence
15	83	34.6	186	6 AR117066	AR117066 Sequence
16	83	34.6	186	6 T06482	T06482 Sequence 19
17	83	34.6	186	6 T0524	T0524 Sequence 62
18	83	34.6	287	3 O05420010	AJ420010 Osterlagl
19	83	34.6	3612	10 MNCOLA4	X067771 Mouse mRNA
20	83	34.6	6512	6 AX306181	AX306181 Sequence
21	83	34.6	6512	10 MUSCOLIA4A	AJ0694 Mus musculu
22	83	34.6	36724	1 SC2G18	AL390188 Streptomy
23	83	34.6	43735	3 CBRG02P14	AC084481 Caenorhab
24	82.5	34.4	437	10 MMA379COL	X91012 M.musculu
25	82.5	34.4	1009	9 BC004412	BC004412 Homo sapi
26	82.5	34.4	1521	10 AF237721	AF237721 Mus muscu
27	82.5	34.4	1605	10 BC030945	BC030945 Mus muscu
28	82.5	34.4	2091	9 BC007574	BC007574 Homo sapi
29	82.5	34.4	2388	10 AF345718	AF345718 Mus muscu
30	82.5	34.4	2437	9 BC011705	BC011705 Homo sapi
31	82.5	34.4	2480	9 HUMCOL3IX	L41162 Homo sapien
32	82.5	34.4	2543	6 AR014074	AR014074 Sequence
33	82.5	34.4	2543	6 AR111836	AR111836 Sequence
34	82.5	34.4	6246	9 HSCOLA3S2	AF026802 Homo sapi
35	82.5	34.4	69252	2 ACT01519	ACT01519 Mus muscu
36	82.5	34.4	160241	9 HS885L7	AL035669 Human DNA
37	82.5	34.4	226854	10 AL669926	AL669926 Mouse DNA
38	82	34.2	426	9 AF230926	AF230926 Macaca mu
39	82	34.2	994	6 AX067354	AX067354 Sequence
40	82	34.2	1485	10 MOSCA1A	M18933 Mouse alpha
41	82	34.2	1881	6 A06090	A06090 Sequence 20
42	82	34.2	1881	6 AR123971	AR123971 Sequence
43	82	34.2	2718	10 BC029697	BC029697 Mus muscu
44	82	34.2	3234	9 HSC3A1R	X15332 Human COL3A
45	82	34.2	3902	9 BC028178	BC028178 Homo sapi

RESULT 1

ALIGNMENTS

HS302D9
LOCUS HS302D9 145880 bp DNA linear PRI 12-DEC-1999
DEFINITION Human DNA sequence from clone RP1-302D9 on chromosome 22 Contains
ACCESSION 282198
VERSION 282198.2 GI:6572207
KEYWORDS HTG.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 145880)
REFERENCE
AUTHORS Bridgeman,A.
JOURNAL Direct Submission
Submitted (08-DEC-1999) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
requests: clonerequest@sanger.ac.uk
On Dec 13, 1999 this sequence version replaced gi:3164067.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence has been finished according to sequence map criteria
as follows. An attempt is made to resolve all sequencing problems,
such as compressions and repeats, but not necessarily within known
annotated human repeat sequence elements (e.g. Alu). Where the
sequence is ambiguous, there is an annotation using the 'unsure'
feature key.
The following abbreviations are used to associate primary accession
numbers given in the feature table with their source databases:
Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMEP; Information
on the WORMEP database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormep This sequence
was generated from part of bacterial clones fromp This sequence
chromosome 22, constructed by the Sanger Centre Chromosome 22
Mapping Group. Further Information can be found at
http://www.sanger.ac.uk/HGP/Chr22
RP1-302D9 is from the library RPCL-1 constructed at the Roswell
Park Cancer Institute by the group of Pieter de Jong. For further
details see http://bacpac.med.buffalo.edu/
VECTOR: pCIPAC2
This sequence is the entire insert of clone RP1-302D9 the true left
end of clone CTA-282P2 is at 69682 in this sequence. The true right
end of clone CTA-415G2 is at 55167 in this sequence.
FEATURES
Source
Location/Qualifiers
1..145880
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="22"
/clone="RP1-302D9"
/clone_1lb="RPCL-1"
188..245
repeat_region
/note="MER3 repeat: matches 144..209 of consensus"
246..571
repeat_region
/note="AluX repeat: matches 1..312 of consensus"
572..759
repeat_region
/note="MER3 repeat: matches 1..144 of consensus"
783..933
repeat_region
/note="MER5A repeat: matches 26..187 of consensus"
1033..1336
repeat_region
/note="AluSP repeat: matches 1..299 of consensus"
1450..1583
repeat_region
/note="MIR repeat: matches 24..160 of consensus"
1687..1752
repeat_region
/note="L2 repeat: matches 2593..2661 of consensus"
2350..2660
repeat_region
/note="AluSc repeat: matches 3..309 of consensus"
2684..2981
repeat_region
/note="AluSq repeat: matches 2..300 of consensus"
3323..3343
repeat_region
/note="MIR repeat: matches 116..136 of consensus"

repeat_region 3344..3652
/note="AluY repeat: matches 1..309 of consensus"
repeat_region 3653..3928
/note="MIR repeat: matches 136..359 of consensus"
repeat_region 3929..4278
/note="MER1B repeat: matches 3..364 of consensus"
4279..4485
repeat_region
/note="MIR repeat: matches 359..568 of consensus"
5073..5176
repeat_region
/note="52 copies 2 mer ct 78 conserved"
5181..5491
repeat_region
/note="AluY repeat: matches 1..311 of consensus"
6369..6485
repeat_region
/note="L2 repeat: matches 2579..2705 of consensus"
6647..6685
repeat_region
/note="MIR repeat: matches 1..23 of consensus"
6686..6987
repeat_region
/note="AluX repeat: matches 1..302 of consensus"
6988..7036
repeat_region
/note="MIR repeat: matches 23..77 of consensus"
7482..7754
repeat_region
/note="AluY repeat: matches 9..290 of consensus"
7775..8060
repeat_region
/note="AluY repeat: matches 1..295 of consensus"
8414..8551
repeat_region
/note="L2 repeat: matches 2553..2706 of consensus"
8914..9030
repeat_region
/note="MIR repeat: matches 147..262 of consensus"
9110..9280
repeat_region
/note="MIR repeat: matches 91..262 of consensus"
9283..9412
repeat_region
/note="MIR repeat: matches 15..144 of consensus"
9521..9679
repeat_region
/note="FAM repeat: matches 3..161 of consensus"
9820..10225
repeat_region
/note="MIR repeat: matches 2..425 of consensus"
complement(10179..10678)
misc_feature
/note="match: GSS: Em:B56592"
complement(10204..10728)
misc_feature
/note="match: GSS: Em:A0701486"
complement(10249..10706)
misc_feature
/note="match: GSS: Em:A0225495"
10312..10383
repeat_region
/note="MIR repeat: matches 79..150 of consensus"
10718..11310
misc_feature
/note="match: GSS: Em:B14024"
10784..11201
misc_feature
/note="match: GSS: Em:B43656"
11838..11946
repeat_region
/note="MIR repeat: matches 20..137 of consensus"
12174..12445
repeat_region
/note="L2 repeat: matches 1988..2275 of consensus"
12444..12642
repeat_region
/note="MIR repeat: matches 63..241 of consensus"
13017..13369
misc_feature
/note="match: STS: Em:G49301"
13331..13397
repeat_region
/note="MIR repeat: matches 174..244 of consensus"
13398..13698
repeat_region
/note="AluSP repeat: matches 1..302 of consensus"
13699..13810
repeat_region
/note="MIR repeat: matches 76..174 of consensus"
13806..13919
repeat_region
/note="MIR repeat: matches 77..189 of consensus"
13945..14060
repeat_region
/note="MIR repeat: matches 24..142 of consensus"
14061..14367
repeat_region
/note="AluY repeat: matches 1..301 of consensus"
14368..14452
repeat_region
/note="MIR repeat: matches 141..225 of consensus"
14589..14679
repeat_region
/note="MIR repeat: matches 173..262 of consensus"
14597..15201
misc_feature


```

misc_feature      14616..15060
                   /note="match: GSS: Em:A0553482"
repeat_region     14868..15040
                   /note="match: GSS: Em:AQ370601"
repeat_region     15071..15188
                   /note="MIR repeat: matches 49. .233 of consensus"
repeat_region     15304..15399
                   /note="L2 repeat: matches 2112. .2239 of consensus"
repeat_region     15490..15662
                   /note="LTRB repeat: matches 1. .99 of consensus"
repeat_region     15669..15727
                   /note="Alusg1 repeat: matches 2. .114 of consensus"
repeat_region     15728..16027
                   /note="LTRB repeat: matches 119. .178 of consensus"
repeat_region     16028..16245
                   /note="Alusc repeat: matches 1. .299 of consensus"
repeat_region     16546..16854
                   /note="LTRB repeat: matches 178. .390 of consensus"
repeat_region     18296..18323
                   /note="Aluy repeat: matches 1. .300 of consensus"
repeat_region     18324..18392
                   /note="MSTR repeat: matches 2. .29 of consensus"
repeat_region     18393..18712
                   /note="MER6-internal repeat: matches 4919. .4993 of
consensus"
repeat_region     18713..19133
                   /note="Aludb repeat: matches 1. .311 of consensus"
repeat_region     19537..20290
                   /note="MER6-internal repeat: matches 4548. .4919 of
consensus"
misc_feature      complement(18872..19230)
                   /note="match: GSS: Em:AQ005063"
misc_feature      19251..19719
                   /note="match: GSS: Em:B14179"
repeat_region     19537..20290
                   /note="HERFH21 repeat: matches 4657. .5784 of consensus"
repeat_region     20317..20382
                   /note="33 copies 2 mer ta 68 conserved"
repeat_region     20513..20666
                   /note="77 copies 2 mer tt 70 conserved"
repeat_region     20682..21008
                   /note="Alusg1 repeat: matches 1. .306 of consensus"
repeat_region     21239..21553
                   /note="HUES-P3 repeat: matches 4410. .4713 of consensus"
repeat_region     21882..22254
                   /note="THEIB repeat: matches 1. .364 of consensus"
repeat_region     22302..22537
                   /note="MER6-internal repeat: matches 2186. .2417 of
consensus"
repeat_region     22538..22850
                   /note="Alusd repeat: matches 1. .313 of consensus"
repeat_region     22851..23801
                   /note="MER6-internal repeat: matches 1210. .2186 of
consensus"
repeat_region     23905..23989
                   /note="MER6-internal repeat: matches 3017. .3102 of
consensus"

Alignment Scores:
Pred. No.:      3.12e-07      Length:      145880
Score:          197.00      Matches:      37
Percent Similarity: 100.00%      Mismatches: 0
Best Local Similarity: 100.00%      Indels: 0
Query Match:    82.08%      Gaps: 0
DB:             9

US-09-997-610-2_COPY_1_45 (1-45) x HS302D9 (1-145880)

QY      1      11eVaVal11eProVal11eThra1aVal11eGluH1sVal1GluVala1aGlyPro 20
         |||||||
Db      34786  ATAGTGCATACCTGCTTATATACGCACATGATGATGAGAGTGTGCTGACT 34845
         |||||||

QY      21      ProAaH1sProAaGProProGluGluVal1GlyProProGlyA1aProGly 37
         |||||||
Db      34846  CCAGCACACCCAGGCCCCAGAGAACTGGGCGCTCTGTCGACACAGT 34896
         |||||||

```


Oy	51683	GGAGGCTCCCAATGATTGGGGCCTCTGTGCACACTGTCCTTCTCGAGCAVCTGTCTCT	51742
Dn	51743	CCTGGGCTCTGTCCTCTGTCCTGTCCTCTGAGCCTCTCT	51781
RESULT 5			
LOCUS	AC011251		
DEFINITION	AC011251.1	173508 bp DNA linear INV 07-JUN-2001	
VERSION	Drosophila melanogaster, chromosome X, region 19F-20A, BAC clone		
ACCESSION	BACR09F10, complete sequence.		
KEYWORDS	AC011251		
SOURCE	AC011251.5 GI:14327739		
ORGANISM	HTG.		
REFERENCE	Drosophila melanogaster.		
AUTHORS	Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.		
TITLE	1 (bases 1 to 173508)		
JOURNAL	Celniker,S.E., Adams,M.D., Kronmiller,B., Tyler,D., Wan,K.H., Holt,R.A., Evans,C.A., Godyne,J.D., Amanosides,P.G., Brandon,R.C., Rogers,Y., An,H., Baldwin,D., Banzon,T., Beeson,K.T., Busam,D.A., Carlson,J.W., Center,A., Champagne,M., Deavenport,L.B., Dietz,S.M., Dodson,K., Dorsett,V., Doup,L.E., Doyle,C., Dresnek,D., Fartan,D., Ferreira,S., Frisbe,E., Galle,R.F., Galle,N.S., George,R.A., Gonzalez,M., Houck,J., Hoskins,R.A., Hostin,D., Howland,T.J., Irbegman,C., Jalali,M., Kruse,D., Li,P., Mattel,B., Moshrefi,A., McIntosh,T.C., Moy,M., Murphy,B., Nelson,C., Nelson,K.A., Nunco,J., Pacled,J., Paragas,V., Park,S., Patel,S., Pfeiffer,B., Phouenavong,S., Pittman,G.S., Puti,V., Richards,S., Scheeler,F., Shantonemong,S., Strong,R., Svitskas,R., Tector,C., Williams,S.M., Zaveril,J.S., Smith,H.O., Rubin,G.M. and Venter,J.C.		
REFERENCE	Sequencing of Drosophila chromosome X, region 19F-20A Unpublished		
AUTHORS	2 (bases 1 to 173508)		
TITLE	Celniker,S.E., Aghayani,A., Arcaina,T.T., Baxter,E., Blazer,J.G., Butenoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L., Doyle,C.M., Fartan,D.E., Galle,R., George,R.A., Harris,N.L., Hoskins,R.A., Houston,K.A., Hummelst,S.R., Karia,K., Kearney,L., Kim,E., Lee,B., Lewis,S., Li,P., Lomton,M.A., Mazda,P., Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacled,J.M., Park,S., Pfeiffer,B., Poon,L., Sequeira,A., Sethi,H., Snir,E., Svitskas,R.K., Wan,K.H., Weinburg,T., Zhang,R., Zierian,L.L. and Rubin,G.M.		
JOURNAL	Direct Submission		
COMMENT	Submitted (05-OCT-1999) Drosophila Genome Center, Lawrence Berkeley Laboratory, MS 64-121, Berkeley, CA 94720, USA On Jun 7, 2001 this sequence version replaced g1:6087507. Sequence submitted by: Berkeley Drosophila Genome Project Lawrence Berkeley National Laboratory, MS 64-121 Berkeley, CA 94720 This sequence was assembled using end sequences from a whole genome shotgun and from subclones of this BAC and its neighboring clones. For further information about this sequence, including its location and relationship to other sequences, please visit our sequence archive Web site (http://www.fruitfly.org/sequence/) or send email to bddgen@fruitfly.berkeley.edu . Location/Qualifiers 1..173508 /organism="Drosophila melanogaster" /strain="y cn bw sp" /db_xref="taxon:7227" /chromosome="X" /map="19F-20A" /clone.lib="BACR09F10 (DL128)" /clone.lib="RPRI-98 (Roswell) Park Cancer Institute Drosophila melanogaster BAC library, partial EcoRI in pBAC3.6")		
FEATURES			
source			
PAGE COUNT	47977 a 37003 c 38363 g 50165 t		
ORIGIN			

Alignment Scores:	1. 12e+03	Length:	173508
Pred. No.:	89.00	Matches:	21
Score:	47.178	Conservative:	4
Percent Similarity:	39.628	Mismatches:	12
Best Local Similarity:	37.088	Indels:	16
Query Match:	3	Gaps:	2
DB:			

US-09-997-610-2_COPY_L_45 (1-45) x AC011251 (1-173508)	Qy	2	valVallieProvalleuiletherAlaValIleGluHisValGluValAla-----	18
Db	91739	GTCCGTCGTCGAAGCGTTGTTGAGGCTCAGTTGTC--GTGCGATGCTTTGGTTGGAT	91795	
Qy	19	-----GlyProProAlaHisProAnp	26	
Db	91736	GGAGGTCCTCCATGGTTGGGGCCCTCTGGTCACCTGGTCTCTCTGGGCATGCTGCTCT	91855	
Qy	27	ProGluGluValGlyProProGlyAlaProGlyLeuPro	39	
Db	91856	CCTGGGCTCTCTGCTCTCTCTGCTCTCTCTGCTCTCTCT	91894	

RESULT 6	292911 bp	DNA	linear	INV 05-OCT-2000
AE003568				
LOCUS				
DEFINITION	292911 bp	DNA	linear	INV 05-OCT-2000
ACCESSION	AE003568			
VERSION	AE003568.2			
KEYWORDS	GI:10726667			
SOURCE	HTG.			
ORGANISM	Drosophila melanogaster			
REFERENCE	Drosophila melanogaster			
AUTHORS	Enkaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Ephydroidea; Drosophila.			
	1 (bases 1 to 292911)			
	Adams,M.D., Celniker,S.E., Holt,R.A., Evans,C.A., Gocayne,J.D.,			
	Maniatis,T., Myers,R.W., J. P.W., Hoskins,R.A., Galle,R.F.,			
	Georgie,R.A., Lewis,S.E., Richards,S., Ashburner,M., Henderson,S.N.,			
	Stratton,R.G., Wortman,J.R., Yandell,M.D., Zhang,Q., Chen,L.X.,			
	Baldwin,C.C., Rogers,Y.H., Blazer,R.G., Chame,M., Pfeiffer,B.D.,			
	Wan,K.H., Doyle,C., Baxter,E.G., Helt,G., Nelson,C.R., Gabor			
	Miklos,G.L., Abiri,J.F., Agbayani,A., An,H.T.,			
	Andrews-Plannkroch,C., Baldwin,D., Ballieu,R.M., Basu,A.,			
	Baxendale,J., Bayraktaroglu,L., Beasley,E.M., Beeson,K.Y.,			
	Benois,P.V., Berman,B.P., Bhandari,D., Bolshakov,S., Borkova,D.,			
	Botchan,M.R., Bouck,J., Brokstein,P., Brothier,P., Burris,K.C.,			
	Bussan,D.A., Butler,H., Cadieu,E., Center,A., Chandra,I.,			
	Cherry,J.M., Cawley,S., Dahlike,C., Davenport,L.B., Davies,P., de			
	Padlos,R., Delcher,A., Deng,Z., Mays,A.D., Dew,I., Dietz,S.M.,			
	Dodson,K., Doup,L.E., Downes,M., Dugan-Rocha,S., Dunikov,B.C.,			
	Dunn,P., Durbin,K.A., Evangelista,C.E., Farz,C., Ferreira,S.,			
	Fleischmann,W., Foster,C., Gabriellian,A., Gatz,N.S.,			
	Gelbart,W.M., Glasser,K., Glodok,A., Gong,F., Gorrell,J.H., Gu,Z.,			
	Guan,P., Harris,M., Harris,N.L., Harvey,D., Helman,T.J.,			
	Henderson,J.R., Houck,U., Hostin,D., Houston,K.A., Howland,T.J.,			
	Wei,M.H., Idegam,C., Jalali,M., Kalush,F., Kapen,G.H., Ke,Z.,			
	Kennison,J.A., Ketchum,K.A., Kimmel,B.E., Kodira,C.D., Kraft,C.,			
	Kravitz,S., Kulp,D., Lai,Z., Laslo,P., Lei,Y., Levitsky,A.A.,			
	Li,J., Li,Z., Liang,Y., Lin,X., Liu,X., Matel,B., McIntosh,T.C.,			
	Moldes,M.P., Moperson,D., Merkulov,G., Milshina,N.V., Mohanty,C.,			
	Morris,D., Moshrefi,A., Mount,S.M., Moy,M., Murphy,B., Murphy,L.,			
	Muzny,D.M., Nelson,D.L., Nelson,D.R., Nelson,K.A., Nixon,K.,			
	Nusskern,D.R., Pacleab,J.M., Palazolo,M., Piltman,G.S., Pan,S.,			
	Pollack,J., Puri,V., Reese,M.G., Reinert,K., Remington,K.,			
	Saunders,R.D., Scheeler,F., Shen,H., Shue,B.C., Sledge-Klamos,I.,			
	Simpson,M., Skupski,M.P., Smith,T., Spier,E., Spirding,A.C.,			
	Stapleton,M., Strong,R., Sun,E., Syvaskas,R., Tector,C., Turner,R.,			
	Venter,E., Wang,A.H., Wang,X., Wang,Z.Y., Wasserman,D.A.,			
	Weinstock,G.M., Weissenbach,J., Williams,S.M., Woodage,T.,			
	Worley,K.C., Wu,D., Yang,S., Yao,Q.A., Ye,J., Yeh,R.F.,			
	Zaveri,J.S., Zhan,M., Zhang,Q., Zhao,Q., Zheng,L., Zheng,X.H.,			


```
/protein_id="AA50837.1"
/db_xref="GI:7295524"
/db_xref="FLYBASE:FBan0001718"
/db_xref="FLYBASE:FBgn0031170"
/translation="MARVTMDKDFLLMKNTLQWNNHKQWVIELVPAIFSLILV
VRTLVDEQGVRYNEQNLTOLNLOKNGSFKEFFILCYSPVNPVAKKIVEAMQS
```

Alignment Scores:

Pred. No.:	1.75e+03	Length:	292911
Score:	89.00	Matches:	21
Percent Similarity:	47.17%	Conservative:	4
Best Local Similarity:	39.62%	Mismatches:	12
Query Match:	37.08%	Indels:	16
OB:	3	Gaps:	2

US-09-997-610-2_COPY_1_45 (1-45) x AE003568 (1-292911)

QY 2 valvalilpvoalleuilehrAlaValillegluhtlsvalgluvalaA-----18

Db 168799 gtctgtgttgAAGTCTGTGTGAGGCTCACTGTCTC---GTGAGTGTGTGTGGTGGAT 168855

QY 19 -----GLYProProAlaHisProArgPro 26

Db 168856 GGAGGTCCCATGTTTGGGGCTCTGTGCTCACCCTGTCTCTCTCTGCGCATCTGTGCTCT 168915

QY 27 ProgluGlValglYProProGlYAlaProGlYLeuPro 39

Ob 168916 CCTGGGCTCTCGTCTCTGTGCTCTCTGCGCTCT 168954

RESULT 7

AC024805/c AC024805 68661 bp DNA linear INV 29-MAY-2002

DEFINITION Caenorhabditis elegans cosmid Y51H7C, complete sequence.

ACCESSION AC024805

VERSION AC024805.1 GI:7140366

KEYWORDS HTG.

SOURCE Caenorhabditis elegans.

ORGANISM Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;

REFERENCE 1 (bases 1 to 68661)

AUTHORS Waterston, R.

TITLE Genome sequence of the nematode C. elegans: a platform for

JOURNAL Science 282 (5396), 2012-2018 (1998)

MEOLINE 99069613

PUBMED 9851916

REFERENCE 2 (bases 1 to 68661)

AUTHORS Bradshaw-Cordum, H. and Oubunue, T.

TITLE The sequence of C. elegans cosmid Y51H7C

JOURNAL Unpublished (2001)

REFERENCE 3 (bases 1 to 68661)

AUTHORS Waterston, R.H.

TITLE Direct Submission

JOURNAL Submitted (01-MAR-2000) Genome Sequencing Center, Washington

REFERENCE 4 (bases 1 to 68661)

AUTHORS Waterston, R.

TITLE Direct Submission

JOURNAL Submitted (24-MAR-2000) Department of Genetics, Washington

REFERENCE 5 (bases 1 to 68661)

AUTHORS University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA

TITLE Waterston, R.

JOURNAL Submitted (24-APR-2001) Department of Genetics, Washington

REFERENCE 6 (bases 1 to 68661)

AUTHORS University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA

TITLE Waterston, R.

JOURNAL Submitted (22-AUG-2001) Department of Genetics, Washington

REFERENCE 7 (bases 1 to 68661)

AUTHORS University, Genome Sequencing Center, 4444 Forest Park Avenue, St.

TITLE Louis, MO 63110, USA

REFERENCE

7 (bases 1 to 68661)

AUTHORS

Waterston, R.
Direct Submission
Submitted (07-SEP-2001) Department of Genetics, Washington
University, Genome Sequencing Center, 4444 Forest Park Avenue, St.
Louis, MO 63110, USA

TITLE

8 (bases 1 to 68661)

JOURNAL

Waterston, R.
Direct Submission
Submitted (29-MAY-2002) Department of Genetics, Washington
University, Genome Sequencing Center, 4444 Forest Park Avenue, St.
Louis, MO 63110, USA

COMMENT

Submitted by:

Genome Sequencing Center
Department of Genetics, Washington University
St. Louis, MO 63110, USA, and
Sanger Centre, Hinxton Hall
Cambridge CB10 1RQ, England
email: rtw@nemadode.wustl.edu and jse@sanger.ac.uk

NOTICE: This sequence may not be the entire insert of this clone. It may be shorter because we only sequence overlapping sections once, or longer because we provide a small overlap between neighboring submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats: all regions were covered by sequence from more than one m13 subclone.

For a graphical representation of this cosmid sequence and its analysis see:
<http://www.wormbase.org/db/seq/sequence?name=Y51H7C;class=Sequence>

NEIGHBORING COSMID INFORMATION

The 5' cosmid is F54D12, 200 bp overlap; the 3' cosmid is H17B01, 200 bp overlap.

NOTES:

Coding sequences below are the result of integration and manual review of the following data: computer analysis using the program Genefinder (P. Green and L. Hillier, personal communication), the large scale EST projects of Yui Kohara (http://www.ddbj.nig.ac.jp/c-elegans/html/CE_INDEX.html) and The C. elegans ORFeome cloning project (<http://wormfdb.dfc.harvard.edu/>), similarity to other proteins from BlastX analyses (<http://blast.wustl.edu/>), sequence conservation with C. briggsae using Jim Kent's WABA alignment program (Genome Research 10:1115-1125, 2000), individual C. elegans GenBank submissions, and personal communications with C. elegans researchers. tRNAs are predicted using the program tRNAscan-SE (Lowe, T.M. and Eddy, S.R., 1997, Nucl. Acids. Res., 25, 955-964).

FEATURES

source

gene

COS

COS

COS

COS

COS

COS

COS

gene

/db_xref="GI:13775524"
/translation="MSLWQIDELGLCDGDIQRKLNINVLCSIGKNNEKMLRRRYV
VOSVEVKETIYENGEGHEDQMDELKP"
1328..1834
/gene="Y51HC.7"
/note="for a graphical representation of this gene see:
<http://www.wormbase.org/db/seq/sequence?name=Y51HC.7>, class
s-Sequence"
join(1328..1450,1505..1630,1682..1834)
/gene="Y51HC.7"
/note="coded for by the following C. elegans cDNAs:
yk560hl1.5, yk118a3.3, yk118a3.5, yk560hl1.3"
/codon_start=1
/product="Hypothetical protein Y51HC.7"
/protein_id="AAK39333.1"
/db_xref="GI:13775525"
/translation="MSSRTIPADVIATLOKFDELTLDLAIVEEVGVGEKHRSNA
HEMALVDTHSMFLNDSLMAVAOATKGCGADKNDDLIDLATFKMTADMKEITNRDPA
PRINKQAANPEFRNLWEQPEOGESSKKAAP"
2368..7728
/gene="Y51HC.6"
/note="for a graphical representation of this gene see:
<http://www.wormbase.org/db/seq/sequence?name=Y51HC.6&class=s-Sequence>"
join(2368..2797,2854..2979,3032..3261,4332..4496,
4570..5048,5864..6340,6802..7129,7558..7728)
/gene="Y51HC.6"
/note="coded for by the following C. elegans cDNAs:
yk402e5.3, yk22a2.3, yk83f8.3, yk20f2.3, yk20f1.3,
yk56qg8.3, yk667a5.3, yk22a2.5, yk73d3.5, yk92t2.5,
yk14sa23.5, yk18ac5.5, CEESQ78F, yk20sb1.5, yk39ae12.5,
yk400bl1.5, yk402e5.5, yk559g2.3, yk559g2.5, yk619e7.3,
yk667a5.5"
/codon_start=1
/product="Hypothetical protein Y51HC.6a"
/protein_id="AAK39338.1"
/db_xref="GI:13775530"
/translation="MPEDILHSKLSPAGRKNGVRHGKEPKVEKOPDFTRKIREI
LLEIEIKREERIEKDIAILEENTIDGGEONRSGFLATRLNMHLVENSAGKOLU
TSAKINSLAVDTISGRSALDAVAKTRPVGCGLTAGMDRGDAEGEIDAIRSEDEF
TASOHILHFLLTDQAVPOIRREFKORDAIDRSHSYEVSKAKSELKTLSRLTESQ
KODVAMORFIKMPLIHEDPEGLORSVFLENOKIDKAEENLAIMAGGGDDNRNAV
LVADILFMFEVSVALETISNLPSLEHSGLKRLDLDPMETLQARIDEFFRLHEFDTR
RLSHFNRLVDDVLIHQKKAAREVEDAPPMEDIASICIMANTSAEMVWFYSRIG
KNEVIRSFGSDGDDEENEARERHLRAEKEDKKMOPLLRSRSGTKMOELLNYCL
LEHFMLKSVOYAKIDSVEKPDAGGLTSIVDVAFILKSTRBAAGSNDVCATIN
NATALIDVINGHLROSIOQCVTYSSNSASFATPAQQGKYKKAADAKOEQLALIN
NSAKSELLIEIQKGLITEMAGVAPRDVERKLEHTSTQGLESAKKLASLAKHOVEL
FKSARKNKIKCGADPYOETIDROMQOVDEVYEADHPMEQGLAQIDLELVENEILPFA
DNYQTULITSSEIARQIEOSLACQCFRYSALQDLDERYOICAYLINVGAMSREKV
GUGQIVSLNETVIDEAMEVHWNSKAMTSATSATIRTLPLPERKVALRADFPVIAIK
SITE"

CDS

join(2368..2797,2854..2979,3032..3261,4332..4496,
4570..5103)
/gene="Y51HC.6"
/note="coded for by the following C. elegans cDNAs:
yk400bl.3, yk400bl.5"
/codon_start=1
/product="Hypothetical protein Y51HC.6b"
/protein_id="AAK39339.1"
/db_xref="GI:13775531"
/translation="MPEDILHSKLSPAGRKNGVRHGKEPKVEKOPDFTRKIREI
LLEIEIKREERIEKDIAILEENTIDGGEONRSGFLATRLNMHLVENSAGKOLU
TSAKINSLAVDTISGRSALDAVAKTRPVGCGLTAGMDRGDAEGEIDAIRSEDEF
TASOHILHFLLTDQAVPOIRREFKORDAIDRSHSYEVSKAKSELKTLSRLTESQ
KODVAMORFIKMPLIHEDPEGLORSVFLENOKIDKAEENLAIMAGGGDDNRNAV
LVADILFMFEVSVALETISNLPSLEHSGLKRLDLDPMETLQARIDEFFRLHEFDTR
RLSHFNRLVDDVLIHQKKAAREVEDAPPMEDIASICIMANTSAEMVWFYSRIG
KNEVIRSFGSDGDDEENEARERHLRAEKEDKKMOPLLRSRSGTKMOELLNYCL
LEHFMLKSVOYAKIDSVEKPDAGGLTRCGRDKKYYKRRLPKQKTC"
complement(7941..10213)
/gene="Y51HC.9"
/note="for a graphical representation of this gene see:

gene

[illegible]

FEATURES	* 188669	278007: contig of 89339 bp in length.
source	Location/Qualifiers	
	1..278007	
	/organism="Caenorhabditis elegans"	
	/db_xref="taxon:6239"	
	/clone="Y51H7"	
BASE COUNT	87340 a 51392 c 48889 g 90302 t	84 others
ORIGIN		
Alignment Scores:		
Pred. No.:	2.27e+03	Length: 278007
Score:	87.50	Matches: 17
Percent Similarity:	58.82%	Conservative: 3
Best local Similarity:	50.00%	Mismatches: 9
Query Match:	36.46%	Indels: 5
DB:	2	Gaps: 1
US-09-997-610-2_COPY_1_45 (1-45) x AC006799 (1-278007)		
OY	17	ValAlaGlyProProlAlaHisProArgProProGluValAlaGlyProProGlyAlaProGly 36
Db	66583	GTACTGTGACCTCCAGGGGTCTCTGGACTCCCGGGTACTGGACCTCCTGGGTACT 66524
OY	37	GlyLeuProGlnTyr-----ThrglyGluLe 45
Db	66523	GGACCAACCCGATACCGGTAGAACCCTGCAACTGGCCATTA 66482
RESULT 9		
AX146422		
LOCUS	AX146422	4428 bp DNA linear PAT 31-MAY-2001
DEFINITION	Sequence 3 from Patent WO0134647.	
ACCESSION	AX146422	
VERSION	AX146422.1	GI:14284845
KEYWORDS		
SOURCE		
ORGANISM	cow.	
	Bos taurus	
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
	Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;	
	Bovidae; Bovinae; Bos.	
REFERENCE	1 (bases 1 to 4428)	
AUTHORS	Bell, M.P., Neff, T.B., Polarek, J.W. and Seeley, T.W.	
TITLE	Animal collagens and gelatins	
JOURNAL	Patent: WO 0134647-A 3 17-MAY-2001;	
FEATURES	FTBROGEN, INC. (US)	
source	Location/Qualifiers	
	1..4428	
	/organism="Bos taurus"	
	/db_xref="taxon:9913"	
BASE COUNT	973 a 1185 c 1400 g 870 t	
ORIGIN		
Alignment Scores:		
Pred. No.:	91.6	Length: 4428
Score:	86.00	Matches: 16
Percent Similarity:	61.54%	Conservative: 0
Best local Similarity:	61.54%	Mismatches: 10
Query Match:	35.83%	Indels: 0
DB:	6	Gaps: 0
US-09-997-610-2_COPY_1_45 (1-45) x AX146422 (1-4428)		
OY	18	AlaGlyProProlAlaHisProArgProProGluValAlaGlyProProGlyAlaProGly 37
Db	538	GCTGCTCTCTCTGGCCACCGGACCCCTGGCACATCTGCTGCTGCTGGC 537
OY	38	LeuProGlnTyrThrgly 43
Db	598	GCTCCAGGATACCAAGST 615
RESULT 10		
AX146424		
LOCUS	AX146424	4428 bp DNA linear PAT 31-MAY-2001
DEFINITION	Sequence 5 from Patent WO0134647.	

ACCESSION	AX146424	GI:14284846
VERSION	AX146424.1	
KEYWORDS		
SOURCE		
ORGANISM	Bos taurus	
	Eumariotata; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovidae; Bovinae; Bos.	
REFERENCE	1 (bases 1 to 4428)	
AUTHORS	Ball M.P., Neff T.B., Polarek J.W. and Seeley T.W.	
TITLE	Animal collagens and gelatins	
JOURNAL	Patent: WO 0134647-A 5 17-MAY-2001;	
FEATURES	FEATURES	
source	Location/Qualifiers	
	1..4428	
	/organism="Bos taurus"	
	/db_xref="taxon:9913"	
BASE COUNT	974 a 1185 c 1399 g 870 t	
ORIGIN		
Alignment Scores:		
Pred. No.:	91.6	Length: 4428
Score:	86.00	Matches: 16
Percent Similarity:	61.54%	Conservative: 0
Best Local Similarity:	61.54%	Mismatches: 10
Query Match:	35.83%	Indels: 0
DB:	6	Gaps: 0
US-09-957-610-2_COPY_1_45 (1-45) x AX146424 (1-4428)		
QY	18 ALaaglyPRoPRoalAhSPRoARgPRoPRoGluGluValGlyPRoPRoGlyAlaProgly 37	
Db	538 GCTGCTCCTCTG6CCACCCG6AGCCCTG6CACACTG6CATCTG6TGGCCCTG6C 597	
QY	38 LeuPRoGlnTyThrely 43	
Db	598 GCTCCAGATACCAAGT 615	
RESULT 11		
LOCUS	AY052763	5358 bp MRNA linear VRF 17-JUN-2002
DEFINITION	Xenopus laevis type XVII collagen short variant mRNA, complete cds.	
ACCESSION	AY052763	
VERSION	AY052763.1	GI:21434788
KEYWORDS		
SOURCE		
ORGANISM	African clawed frog.	
	Xenopus laevis	
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Amphibia; Batrachia; Anura; Mesobatrachia; Pipridae; Pipridae; Xenopodidae; Xenopus.	
REFERENCE	1 (bases 1 to 5358)	
AUTHORS	Clamag, H., Peterson, J., Pihlajaniemi, T. and Destree, O.	
TITLE	Cloning of three variants of type XVII collagen and their expression patterns during Xenopus laevis development	
JOURNAL	Mech. Dev. 114 (1-2), 109-113 (2002)	
MEDLINE	21993083	
REFERENCE	11996984	
AUTHORS	2 (bases 1 to 5358)	
TITLE	Elamag, H., Peterson, J., Pihlajaniemi, T. and Destree, O.	
JOURNAL	Direct Submission	
	Submitted (28-AUG-2001) Collagen Research Unit, Blocenter and Department of Medical Biochemistry, University of Oulu, P. O. Box 5000, Oulu 90014, Finland	
FEATURES	Location/Qualifiers	
source	1..5358	
	/organism="Xenopus laevis"	
	/db_xref="taxon:8355"	
CDS	1..3924	
	/codon_start=1	
	/product="type XVII collagen short variant"	
	/protein_id="AA114257.1"	
	/db_xref="GI:21434789"	


```
/translation="MILRCTWVLEFLYASVHALVSSQDKDGFTEVGLLOIGEPVNP  
EIRTVFGEDTKPAYVDFKENTGOVAQYHILPNMVRDESLFHPHSTPNKNGILFAIT  
DASQIYVIGKRLSEVNDGKODIIFETPEPSQDSYVATNPVSLNMRSPALFAIT  
DEEVLVMDCELELRRVRERPERSEMELEDSGLFYAAGAGADPDRIOGVINDLITVAN  
PRAEROCEDSDDDGTISDFSGIENLNDLVTSPPVIVVQPIVPPVLPVLPVLPVLPV  
EIVTVGGQOQGIKEKGDHGEKSGRLGPKGDSASVPASDSNGEKQGEVYKGS  
AGFVYGPGRQKGTMTGPRGPPAGPPGPIILOGSDGTSVQVPGPPGPPAGPAGK  
DGEPPDGEDGKTGEVGPGRGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG  
TDKLFITMEGSSGSDLEHYKRGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG  
LPGLPKKEGTGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG  
PGAAGPGLAGLPGRMGRQPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG  
PGLGKAGEONPGLPGLGVKSGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG  
GHDEVGTPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG  
PGTHGAKGEKGTGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG  
PGRGMNGIKGEKGEKASGSGLRGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG  
PPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG  
YPLGPKAGESVGPGRGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG  
ISTGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG  
GFRRIILGELYTLTSGPQDNEVAEOAVVYVPHRNKDSIDTLOSTEEMSNANVSGA  
RPMGDPRLPEPPRVDPDPDHPHGNPRLPEPPVTHHODENPALHLVALNAPISGSM  
KSTIRGVDFGCFEORAKAGLHCTFRAPLSSRIODLXSYVRADROSQVTVNLDEVLTD  
NMDSLFSGSESNQMSGARITISFDKRDVTTDTPQKVMHSDAKGRKLTSTYCEYMR  
TDESAVVGQASLISGKLEORPQSCNKNFTVLCIENSFMTNNRK"
```

BASE COUNT 1403 a 1234 c 1447 g 1274 t
ORIGIN
Alignment Scores:
Pred. No.: 119 Length: 5358
Score: 85.50 Matches: 18
Percent Similarity: 62.50% Conservative: 2
Best local Similarity: 56.25% Mismatches: 7
Query Match: 35.62% Indels: 5
DB: 5 Gaps: 2

US-09-997-610-2_COPY_1_45 (1-45) x AY052763 (1-5358)

Qy 12 ILEGUHSVAGLVALAGLYPROFALHISPROARGPROGLUCLUVALGly 31
:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||

Db 1402 TTGGAGCATGTCAAG-----GACCCCTGTGTTGCCCGGCCCCCA-----GGT 1446

Qy 32 PROPRGGLYALPRGGLYLEUPROGLINTYrthrGly 43
|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||

Db 1447 CCTCCAGGTACTACCTGCTTACCCGAGTACCCGGA 1482

RESULT 12

AC079422 256417 bp DNA linear HTG 01-SEP-2000

LOCUS Mus musculus chromosome 16 clone RP23-109A22, WORKING DRAFT

DEFINITION AC079422

AC079422.1 GI:9958034

AC079422.1 GI:9958034

HTG: HTGS_PHASE1; HTGS_DRAFT.

KEYWORDS Mus musculus.

SOURCE Mus musculus.

ORANISM Mus musculus.

REFERENCE 1 (bases 1 to 256417)

AUTHORS DOE Joint Genome Institute.

TITLE Unpublished

JOURNAL 2 (bases 1 to 256417)

REFERENCE DOE Joint Genome Institute.

AUTHORS Direct Submission

TITLE Submitted (01-SEP-2000) Production Sequencing Facility, DOE Joint

JOURNAL Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA

COMMENT -----Genome Center

Center: Joint Genome Institute

Center Code: JGI

Web site: http://www.jgi.doe.gov

Project Information

Center Project Name: 0

Center clone name: RPCI-23_109A22

Summary Statistics

Consensus quality: 161724 bases at least Q40
Consensus quality: 191966 bases at least Q30
Consensus quality: 205448 bases at least Q20
Estimated insert size: 220300; agarose-tp estimation
Estimated insert size: 250317; sum-of-contigs estimation
Quality coverage: 3.11 in Q20 bases; agarose-tp estimation
Quality coverage: 2.74 in Q20 bases; sum-of-contigs estimation.
NOTE: This is a "working draft" sequence. It currently
* consists of 62 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1	1101: contig of 1101 bp in length
1102	1201: gap of unknown length
1202	2412: contig of 1211 bp in length
2413	2512: gap of unknown length
2513	3675: contig of 1163 bp in length
3676	3775: gap of unknown length
3776	4879: contig of 1104 bp in length
4880	4979: gap of unknown length
4980	6106: contig of 1127 bp in length
6107	6206: gap of unknown length
6207	7322: contig of 1116 bp in length
7323	7422: gap of unknown length
7423	8556: contig of 1134 bp in length
8557	8656: gap of unknown length
8657	9838: contig of 1182 bp in length
9839	9938: gap of unknown length
9939	11058: contig of 1120 bp in length
11059	11158: gap of unknown length
11159	12325: contig of 1167 bp in length
12326	12425: gap of unknown length
12426	13646: contig of 1221 bp in length
13647	13746: gap of unknown length
13747	14871: contig of 1125 bp in length
14872	14971: gap of unknown length
14972	16434: contig of 1463 bp in length
16435	16534: gap of unknown length
16535	17812: contig of 1278 bp in length
17813	17912: gap of unknown length
17913	19230: contig of 1318 bp in length
19231	19330: gap of unknown length
19331	20400: contig of 1070 bp in length
20401	20500: gap of unknown length
20501	22256: contig of 1756 bp in length
22257	22356: gap of unknown length
22357	23977: contig of 1621 bp in length
23978	24077: gap of unknown length
24078	25281: contig of 1204 bp in length
25282	25381: gap of unknown length
25382	27415: contig of 2034 bp in length
27416	27515: gap of unknown length
27516	29277: contig of 1762 bp in length
29278	29377: gap of unknown length
29378	32364: contig of 2987 bp in length
32365	32464: gap of unknown length
32465	33843: contig of 1379 bp in length
33844	33943: gap of unknown length
33944	36718: contig of 2775 bp in length
36719	36818: gap of unknown length
36819	38355: contig of 1537 bp in length
38356	38455: gap of unknown length
38456	39544: contig of 1089 bp in length
39545	39644: gap of unknown length
39645	42096: contig of 2452 bp in length
42097	42196: gap of unknown length
42197	44438: contig of 2242 bp in length
44439	44538: gap of unknown length

44539	contig of 2112 bp in length
44651	gap of unknown length
46751	contig of 2186 bp in length
49037	gap of unknown length
52140	contig of 3104 bp in length
52141	gap of unknown length
52241	contig of 2507 bp in length
54748	gap of unknown length
55848	contig of 4170 bp in length
55017	contig of unknown length
55018	gap of unknown length
59118	contig of 3684 bp in length
62801	gap of unknown length
62901	contig of 3616 bp in length
65912	gap of unknown length
66517	contig of 3657 bp in length
70275	gap of unknown length
70375	contig of 5165 bp in length
75540	gap of unknown length
75640	contig of 3670 bp in length
79310	gap of unknown length
79410	contig of 5280 bp in length
84690	gap of unknown length
84790	contig of 5021 bp in length
88811	gap of unknown length
89911	contig of 5533 bp in length
95503	gap of unknown length
95604	contig of 5325 bp in length
100928	gap of unknown length
101029	contig of 4485 bp in length
105514	gap of unknown length
105614	contig of 4776 bp in length
110390	gap of unknown length
110489	contig of 4363 bp in length
114853	gap of unknown length
114953	contig of 5774 bp in length
120726	gap of unknown length
120826	contig of 6233 bp in length
127059	gap of unknown length
127060	contig of 6448 bp in length
127153	gap of unknown length
133607	contig of 5921 bp in length
133708	gap of unknown length
133928	contig of 8057 bp in length
139629	gap of unknown length
139729	contig of 6976 bp in length
147785	gap of unknown length
147885	contig of 6021 bp in length
147886	gap of unknown length
154861	contig of 6694 bp in length
154961	gap of unknown length
154962	contig of 6021 bp in length
160982	gap of unknown length
161083	contig of 6694 bp in length
161776	gap of unknown length
167877	contig of 7247 bp in length
175123	gap of unknown length
175124	contig of 8241 bp in length
175223	gap of unknown length
183464	contig of 7816 bp in length
183465	gap of unknown length
191380	contig of 9452 bp in length
191381	gap of unknown length
191481	contig of 9452 bp in length
200932	gap of unknown length
201032	contig of 6555 bp in length
201033	gap of unknown length
207587	contig of 9280 bp in length
207588	gap of unknown length
216967	contig of 12519 bp in length
216968	gap of unknown length
229647	contig of 11746 bp in length
229747	gap of unknown length
241493	contig of 11746 bp in length
241593	gap of unknown length
256417	contig of 14825 bp in length
Location/Qualifiers	

[illegible]

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: February 19, 2003, 14:57:18 ; Search time 2076.08 seconds
(without alignments)
2088.706 Million celli updates/sec

Title: US-09-997-610-2_COPY_1_149
Perfect score: 790
Sequence: 1 IVVLPVLTAVIEHVEVAGP.....KEEISKQOSIQEVTWVLLKA 149

Scoring table:
BLOSUM62
Xgapop 10.0 , Ygapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2054640 segs, 14551402878 residues
Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL=frame_plus_p2n_model -DEV=xlh
-O=/cgn2.1/USPIO_Spool/US09997610/runat_10022003.160823.23819/app.query.fasta_1.1635
-DB=genEmbl -QFMT=fastap -SUFFR=rge -MINMATCH=0.1 -LOOFC=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blsum62 -TRANS=human40.cdi -LIST=45
-DOCALLIN=200 -THR.SCORE=PCT -THR.MAX=100 -THR.MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=plc -NORM=ext -HEADSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09997610.BCGN.1.13965.@runat.10022003.160823.23819 -MCP=6 -ICPU=3
-NO_XLPHY -NO_MMAP -LARGEQUERY -NEG.SCORES=0 -WAIT -LONGLOG -DEV.TIMEOUT=120
-WARN.TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -Fgapop=6 -Fgapext=7
-YGAPOP=10 -YGAPEXT=0.5 -DELop=6 -DELEXT=7

Database :
1: genEmbl:*
2: gb_ba:*
3: gb_hhg:*
4: gb_in:*
5: gb_om:*
6: gb_ov:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*

29: em_vl:*
30: em_hhg_hum:*
31: em_hhg_inv:*
32: em_hhg_other:*
33: em_hhg_mus:*
34: em_hhg_pin:*
35: em_hhg_rtd:*
36: em_hhg_mam:*
37: em_hhg_vrt:*
38: em_sy:*
39: em_hhggo_hum:*
40: em_hhggo_mus:*
41: em_hhggo_other:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	623	78.9	145880	9 HS302D9	282198 Human DNA s
2	330	41.8	1265	10 TMSHP20A	D12974 Tamias asia
3	291.5	36.9	5121	10 AB067779	AB067779 Tamias si
4	220	27.8	2005	10 TMSHP25	D12975 Tamias asia
5	169.5	21.5	1918	10 RNO131848	AJ131848 Rattus no
6	167.5	21.2	1385	10 TMSHP27	D12976 Tamias asia
7	167.5	21.2	2609	5 CHKX	M13496 Chicken typ
8	167	21.1	559	10 AB067813S3	AB067813 Callioscu
9	163.5	20.7	2235	6 AX332258	AX332258 Sequence
10	163.5	20.7	2235	9 HSCOL8A1	X57527 Human COL8A
11	163.5	20.7	2415	4 RABCOLV111	J05042 Rabbid type
12	162.5	20.6	2506	9 BC013581	BC013581 Homo sapi
13	162.5	20.6	117000	9 AC069222	AC069222 Homo sapi
14	162.5	20.6	145880	9 HS302D9	282198 Human DNA s
15	162.5	20.6	152354	2 AC120598	AC120598 Homo sapi
16	162.5	20.6	166992	2 AC022883	AC022883 Homo sapi
17	162.5	20.6	182978	30 AC067824	AC067824 Homo sapi
18	161	20.4	7089	4 AF222861	AF222861 Sus scro
19	159.5	20.2	1986	10 NMCOL8A	X53556 Bovine COL1
20	159	20.1	3143	4 BTCOL10A1	Z21610 M. musculus
21	158	20.0	3422	10 MMA1TXCOL	X67348 Mus musculu
22	158	20.0	9331	10 MMCOL10A	AC119229 Mus muscu
23	158	20.0	68834	2 AC119229	AC016087 Homo sapi
24	158	20.0	179303	2 AC016087	AC021709 Mus muscu
25	157.5	19.9	206213	2 BC021709	BC021709 Mus muscu
26	157.5	19.9	2435	10 BC011061	X63013 M. musculus
27	156	19.7	731	10 MMA1XCOL	X65120 H. sapiens C
28	156	19.7	1973	9 HS10A1COL	X72580 Homo sapien
29	156	19.7	3215	9 HSCOLX3	AX333243 Sequence
30	156	19.7	3226	6 AX333243	X60382 H. sapiens C
31	156	19.7	3226	6 HSCOLX1X	G28608 human STS S
32	156	19.7	3226	11 G28608	X98568 H. sapiens t
33	156	19.7	10058	9 HSCOLX	AL121963 Human DNA
34	156	19.7	107553	9 HS1336014	AL353573 Homo sapi
35	156	19.7	582	4 AF155373	AF155373 Canis fam
36	155.5	19.7	582	4 AF17206	AF17206 Canis fam
37	152.5	19.3	947	10 BC028770	BC028770 Bos tauru
38	152.5	19.3	1134	4 AF269230	AF269230 Bos tauru
39	152.5	19.3	200942	2 AC125396	AC125396 Mus muscu
40	151	19.1	4443	2 AK074129	AK074129 Homo sapi
41	151	19.1	4908	6 AX430953	AX430953 Sequence
42	151	19.1	128133	2 AL138787	AL138787 Homo DNA
43	151	19.1	173817	2 AC012141	AC012141 Homo sapi
44	150.5	19.1	1152	6 AX195207	AX195207 Sequence
45	150.5	19.1	1152	6 AX358517	AX358517 Sequence

RESULT 1

ALIGNMENTS

HS302D9 145880 bp DNA linear PRI 12-DEC-1999
LOCUS Human DNA sequence from clone RPI-302D9 on chromosome 22 contains
DEFINITION
ACCESSION Z82198
VERSION Z82198.2 GI:6572207
KEYWORDS HTG.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE 1 (bases 1 to 145880)
AUTHORS Bridgeman, A.
TITLE Direct submission
JOURNAL Submitted (08-DEC-1999) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
COMMENT
requests: clonerequest@sanger.ac.uk
On Dec 13, 1999 this sequence version replaced gi:3164067.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
correlation annotation may not be found in the sequence submission
only a small overlap as described above.
This sequence has been finished according to sequence map criteria
as follows. An attempt is made to resolve all sequencing problems,
such as compressions and repeats, but not necessarily within known
annotated human repeat sequence elements (e.g. Alu). Where the
sequence is ambiguous, there is an annotation using the 'unsure'
feature key.
The following abbreviations are used to associate primary accession
numbers given in the feature table with their source databases:
Em: EMBL; SW: SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information
on the WORMPEP database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep
This sequence
was generated from part of bacterial clone contigs of human
chromosome 22, constructed by the Sanger Centre Chromosome 22
Mapping Group. Further information can be found at
http://www.sanger.ac.uk/MGP/Chr22
RPI-302D9 is from the library RPCI-1 constructed at the Roswell
Park Cancer Institute by the group of Pieter de Jong. For further
details see http://baopac.med.buffalo.edu/
VECTOR: pcypAC2
This sequence is the entire insert of clone RPI-302D9 The true left
end of clone CTA-282F2 is at 69682 in this sequence. The true right
end of clone CTA-415G2 is at 55167 in this sequence.
FEATURES
source
1..145880
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="22"
/clone="RPI-302D9"
/clone_lib="RPCI-1"
188..245
repeat_region
/note="MER3 repeat: matches 144..209 of consensus"
246..571
repeat_region
/note="AluX repeat: matches 1..312 of consensus"
572..759
repeat_region
/note="MER3 repeat: matches 1..144 of consensus"
783..933
repeat_region
/note="MER3A repeat: matches 26..187 of consensus"
1033..1336
repeat_region
/note="AluSP repeat: matches 1..299 of consensus"
1450..1583
repeat_region
/note="MIR repeat: matches 24..160 of consensus"
1687..1752
repeat_region
/note="L2 repeat: matches 2593..2661 of consensus"
2350..2660
repeat_region
/note="AluSC repeat: matches 3..309 of consensus"
2684..2981
repeat_region
/note="AluJ repeat: matches 2..300 of consensus"
3323..3343
repeat_region
/note="MIR repeat: matches 116..136 of consensus"

repeat_region
3344..3652
/note="AluY repeat: matches 1..309 of consensus"
repeat_region
3653..3928
/note="MIR repeat: matches 136..359 of consensus"
repeat_region
3929..4278
/note="THE1B repeat: matches 3..364 of consensus"
4279..4485
repeat_region
/note="MIR repeat: matches 359..568 of consensus"
5073..5176
repeat_region
/note="52 copies 2 mer ct 78 conserved"
5181..5491
repeat_region
/note="AluB repeat: matches 1..311 of consensus"
6369..6485
repeat_region
/note="L2 repeat: matches 2579..2705 of consensus"
6647..6685
repeat_region
/note="MIR repeat: matches 1..23 of consensus"
6686..6987
repeat_region
/note="AluX repeat: matches 1..302 of consensus"
6988..7036
repeat_region
/note="MIR repeat: matches 23..77 of consensus"
7482..7754
repeat_region
/note="AluB repeat: matches 9..290 of consensus"
7775..8060
repeat_region
/note="AluY repeat: matches 1..295 of consensus"
8414..8551
repeat_region
/note="L2 repeat: matches 2553..2706 of consensus"
8914..9030
repeat_region
/note="MIR repeat: matches 147..262 of consensus"
9110..9280
repeat_region
/note="MIR repeat: matches 91..262 of consensus"
9283..9412
repeat_region
/note="MIR repeat: matches 15..144 of consensus"
9521..9679
repeat_region
/note="FAM repeat: matches 3..161 of consensus"
9820..10225
repeat_region
/note="MIR repeat: matches 2..425 of consensus"
10179..10678
repeat_region
/note="match: GSS: Em:B56592"
10204..10720
repeat_region
/note="match: GSS: Em:AQ701486"
10249..10706
repeat_region
/note="match: GSS: Em:AQ225495"
10312..10383
repeat_region
/note="MIR repeat: matches 79..150 of consensus"
10718..11310
repeat_region
/note="match: GSS: Em:B14024"
10784..11201
repeat_region
/note="match: GSS: Em:B43656"
11838..11946
repeat_region
/note="MIR repeat: matches 20..137 of consensus"
12174..12445
repeat_region
/note="L2 repeat: matches 1988..2275 of consensus"
12444..12642
repeat_region
/note="MIR repeat: matches 63..241 of consensus"
13017..13369
repeat_region
/note="match: STS: Em:G49301"
13331..13397
repeat_region
/note="MIR repeat: matches 174..244 of consensus"
13398..13698
repeat_region
/note="AluX repeat: matches 1..302 of consensus"
13699..13810
repeat_region
/note="MIR repeat: matches 76..174 of consensus"
13806..13919
repeat_region
/note="MIR repeat: matches 77..189 of consensus"
13945..14060
repeat_region
/note="MIR repeat: matches 24..142 of consensus"
14061..14367
repeat_region
/note="AluY repeat: matches 1..301 of consensus"
14368..14452
repeat_region
/note="MIR repeat: matches 141..225 of consensus"
14589..14679
repeat_region
/note="MIR repeat: matches 173..262 of consensus"
14597..15201
misc_feature

	misc_feature	/note="match: GSS: Em:AQ553482" 14616..15060 /note="match: GSS: Em:AQ370601"
	repeat_region	14868..15040 /note="MIR repeat: matches 49..233 of consensus"
	repeat_region	15071..15188 /note="L2 repeat: matches 2112..2239 of consensus"
	repeat_region	15304..15399 /note="MTIB repeat: matches 1..99 of consensus"
	repeat_region	15490..15662 /note="AlusJ1 repeat: matches 2..114 of consensus"
	repeat_region	15669..15737 /note="MTIB repeat: matches 119..178 of consensus"
	repeat_region	15728..16027 /note="Alusc repeat: matches 1..299 of consensus"
	repeat_region	16028..16245 /note="MTIB repeat: matches 178..390 of consensus"
	repeat_region	16546..16854 /note="Alur repeat: matches 1..300 of consensus"
	repeat_region	18296..18323 /note="MSTA repeat: matches 2..29 of consensus"
	repeat_region	18324..18392 /note="MER6-internal repeat: matches 4919..4993 of consensus"
	repeat_region	18393..18712 /note="Alub repeat: matches 1..311 of consensus"
	repeat_region	18713..19133 /note="MER6-internal repeat: matches 4548..4919 of consensus"
	misc_feature	complement(18872..19230) /note="match: GSS: Em:AQ005063"
	misc_feature	19251..19719 /note="match: GSS: Em:B14179"
	repeat_region	19537..20290 /note="HERVH21 repeat: matches 4657..5784 of consensus"
	repeat_region	20317..20382 /note="33 copies 2 mer ta 68 conserved"
	repeat_region	20513..20666 /note="77 copies 2 mer tt 70 conserved"
	repeat_region	20682..21008 /note="AlusJ1 repeat: matches 1..306 of consensus"
	repeat_region	21239..21553 /note="HUER-P3 repeat: matches 4410..4713 of consensus"
	repeat_region	21882..22254 /note="THELB repeat: matches 1..364 of consensus"
	repeat_region	22302..22537 /note="MER6-internal repeat: matches 2186..2417 of consensus"
	repeat_region	22538..22850 /note="Alusp repeat: matches 1..313 of consensus"
	repeat_region	22851..23801 /note="MER6-internal repeat: matches 1210..2186 of consensus"
	repeat_region	23905..23989 /note="MER6-internal repeat: matches 3017..3102 of consensus"
	Alignment Scores:	
	Pred. No. :	1 4e-44 Length: 145880
	Score:	623.00 Matches: 117
	Percent Similarity:	100.00% Conservatvie: 0
	Best Local Similarity:	100.00% Mismatches: 0
	Query Match:	78.86% Indels: 0
	DB:	9 Gaps: 0
US-09-997-610-2.COPY_1.L149 (1-149) * HS30209 (1-145880)		
OY	33 ProglyAlaProGlyLeuProGlnTyrThrGlyGluIleSerGluMetTrpLysCysPro	52
Dp	37262 CCGTGAGCCCAAGCTTTACACAAATPACACGACGAATAAAGTAAGACAATAATGCCCC	37321
OY	53 CysProAspIleGluArgSerAlaIleThrValLysLeuSerGlyLysLeuProLeuPro	72
Db	37322 TGTTCGGAATAGAAAAGGCACGGCTTTACTAGTCAAGCTCAGTGGAAAACTTCCCTTCCT	37381

QY	73	PheLysProIleIleIlePheThrGlyValLeuTYrAsnAlaGlnArgAspLeuLysGluAla	92
Db	37382	TTCAAGGCCCATCATCTTTCACAGGGGCTCGTCAATGCCACAGGAGATTTAAGAGGCC	37441
QY	93	MetGlyValPheAlaCysArgValProGlyAsnTYrTyrSerSerPheAspValGluLeu	112
Db	37442	ATGGCAGCTTTTGCTTGCCAGGGTCCCTGGGAATTACTCTCAAGCTTTGATGTTAGCTG	37501
QY	113	HisHisCysLysValAsnIleTrrPheuneTarGlySgInIleLeuAlaSnLysGluGln	132
Db	37502	CATCATTCGCAAGGTGAATATTGGCTAATGAGCAAGCAAAATTTGGCTAATTAAGACAGA	37561
QY	133	IleSerLysGlnGlnSerIleGlnGluValIleTrrPvalLeuLeuAla	149
Db	37562	ATTTCYAAACAGCAAGCAATTCAGAGGCTGGCTGGCTGTGTAAAGGCA	37612
RESULT 2			
LOCUS	TMSHP20A	1265 bp	mRNA
DEFINITION	Tamias asiaticus mRNA for Hp-20, complete cds.	linear	ROD 03-FEB-1999
ACCESSION	D12974		
VERSION	D12974.1	GI:287467	
KEYWORDS	HP-20; collagen-like domain; hibernation-related protein; plasma protein.		
SOURCE	Tamias asiaticus liver cDNA to mRNA, clone:pcmw20-7.		
ORGANISM	Tamias sibiricus		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Sciuridae; Sciurinae; Tamias.		
AUTHORS	1 (bases 1 to 1265)		
TITLE	Takamatsu,N., Ohba,K., Kondo,J., Kondo,N. and Shiba,T.		
JOURNAL	Hibernation-associated gene regulation of plasma proteins with a collagen-like domain in mammalian hibernators		
MEDLINE	Mol. Cell. Biol. 13 (3), 1516-1521 (1993)		
REFERENCE	93180798		
AUTHORS	2 (bases 1 to 1265)		
TITLE	Takamatsu,N.		
JOURNAL	Direct Submission		
REFERENCE	Submitted (21-AUG-1992) Nobuhiko Takamatsu, School of Hygienic Sciences, Kitasato University, Laboratory of Molecular Biology; 1-15-1 Kitasato, Sagamihara, Kanagawa 228, Japan (Tel:0427-78-9408, Fax:0427-78-9403)		
FEATURES			
source	Location/Qualifiers		
	1..1265		
	/organism="Tamias sibiricus"		
	/db_xref="taxon:64680"		
	/clone="pcmw20-7"		
	/tissue_type="liver"		
	89..679		
	/note="collagen-like domain at nt 161-277"		
	/codon_start=1		
	/product="HP-20"		
	/protein_id="BAA02351.1"		
	/db_xref="GI:287468"		
	/translation="MTDYLRLAIFVLWNLVDVSCSGPPGVPYGVGPGVPPRP		
	PGCGCAARPCDPPGKPGSVKCPGCRSAPFYKRSGRLLPPSEPVEYTVELYNTRDL		
	KESLGVNCEBGNHFSFDVETLCKIKLMLNHLIOWEKKHOLSKNEYENASGAMT		
	MLPLQKGVLEADVETEEDQAVVIFYFSGFLIS"		
	89..157		
	158..676		
	/product="HP-20"		
	1238..1243		
	polya_signal		
	polya_site		
	1265		
BASE COUNT	354 a 304 c 288 g 319 t		
ORIGIN			
Alignment Scores:			
Pred. No.:	1..34e-20	Length:	1265
Score:	330.00	Matches:	69
Percent Similarity:	61.22%	Conservative:	21
Best Local Similarity:	46.94%	Mismatch:	45
Query Match:	41.77%	Indels:	12


```

Db      481 GTAACGAGCATCCGCTGCTTCCTTACTGTCTTCCAAAGCTTACCGAGCGGTA 540
Oy      73 PhelysProIIlePhethrGlyValleuTyraSnaIagInarGaSpLeuIySgluaIa 92
         ||||| ||| ::||| ||||| |||||
Db      541 GGTCGTCCTCATTCATTCATGAGATTCTGTACAACGACGACGACCTTGACCCACAAG 600
Oy      93 MetGlyValPheAlaCysArgValProGlySantYrYrSerSerPheAspValIgluLeu 112
         ||||| ||| ::||| ||||| |||||
Db      601 TCTGAACTCTTACCTTAAGATCCAGCATATCTATTTCTTCAATCATTGATG 660
Oy      113 HIsHIScYsLySValaSnIlePleu-MetArgLySglInleuAlaSnIySgluI 132
         ||| ::| ||| ::|||
Db      661 AAGGGACATCAGCTTGGGTAGCCCTGTATAAGATGGCA-----CGCCACGA 708
Oy      132 uIleSerLySglInSerIleGIngluValThrTrpValleuLeuIyS 148
         : ::::| ||| ::| |||||::| ||::|
Db      709 ATGTACACGTATGATGAGTACAGCAAGGCTACTTGATCAGGCTTCAG 757

RESULT 6
TMSHP27 1385 bp mRNA linear ROD 03-FEB-1999
LOCUS TMSHP27 Tamias asiaticus mRNA for Hp-27, complete cds.
DEFINITION D12976
ACCESSION D12976
VERSION D12976.1 GI:287471
KEYWORDS HP-27; collagen-like domain; hibernation-related protein; plasma protein.
SOURCE Tamias asiaticus liver cDNA to mRNA, clone:pcM27-3.
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Rodentia; Sciurognathi; Sciuridae; Sciurinae;
          Tamias.
REFERENCE 1 (bases 1 to 1385)
AUTHORS Takamatsu, N., Ohba, K., Kondo, J., Kondo, N. and Shiba, T.
TITLE Hibernation-associated gene regulation of plasma proteins with a
JOURNAL Mol. Cell. Biol. 13 (3), 1516-1521 (1993)
MEDLINE 93180798
REFERENCE 2 (bases 1 to 1385)
AUTHORS Takamatsu, N.
TITLE Direct Submission
JOURNAL Submitted (21-AUG-1992) Nobuhiko Takamatsu, School of Hygienic
          Sciences, Kitasato University, Laboratory of Molecular Biology,
          1-15-1 Kitasato, Sagamihara, Kanagawa 228, Japan (tel:0427-78-9408,
          Fax:0427-78-9403)

FEATURES
source 1..1385
         /organism="Tamias sibiricus"
         /db_xref="taxon:64680"
         /clone="pcM27-3"
         /tissue_type="liver"
         65..712
         /note="collagen-like domain at nt 191-307"
         /codon_start=1
         /product="Hp-27"
         /protein_id="BAA02353.1"
         /db_xref="GI:287472"
         /translation="NYEAGKRASPMGAGIIMIALSVLMHVYCSSETGPNPESCAVPGP
         OGPGMRGPGTGGKPGPGMGNPGPLGPGPGIVNVCHSGTSAAFAVKANELPPA
         BSGPYIEKAHDAQGHFDATGCTFCPPVGLVDFGFHIEAVRAKRYSLMNGQVM
         EKEADQDGYEHISGITALIDLGMDRWLENKLSQTDLEKGYVAVSEGLIHEN"
         65..154
         sig_peptide
         mat_peptide
         polyA_signal
         polyA_site 1385..1374
BASE COUNT 439 a 285 c 311 g 350 t
ORIGIN

```

```

Query Match: 21.208 Indels: 29
DB: 10 Gaps: 4
us-09-997-610-2_copy_1_149 (1-149) x TMSHP27 (1-1385)

Oy      6 ValLeuIleThrAlaValIleGIuHISVal-----
Db      116 AATTTGGCCCTGTCTGTACTAATGACGCGTGTGTCTTGAATCTCAAGAAATCTCGAA 175
         ::||| ::|||::| |||||
Oy      16 -----GluValAlaGlyPro-----Pro 21
         ||| |||||
Db      176 TCTGTATATGTTCCAGGACCTCAAGGACCTCCAGGATCGGGGCTCCCGCATACACT 235
         ||| |||||
Oy      22 AlaHISProArTPProGluGluValGlyProProGlyAlaProGlyLeuProGlnTyr 41
         ||| ||||| ||| ||||| |||||
Db      236 GGAACACGAGCCGACACGAGATGGAATGTTTCCAGGATCTACGAGCCACAGGCGCA 295
         ||| |||||
Oy      42 ThrGlyIuIleSerGluMetThrIySyrProGlySyrProAspIleGluArgSerAlaPhe 61
         ||| |||||
Db      296 CCAGCG-----ANGACT--GTGAACCTGCCACAGCAAGAAAGAAATCGGCTTT 340
         ||| |||||
Oy      62 ThrAlaIySleuSerGlyIySleuProLeuProPheIySProIleIlePheThrGlyVal 81
         ||||| ::| ||| ||| ::|||::| |||||
Db      341 GCACTGAAGCCAAATGAGCTGCCCCAGGCTCCCTCCAGCCGATCTTCAAGAAAGCC 400
         ||||| ::| ||| ||| |||||
Oy      82 LeuTyraSnaIagInarGaSpLeuIySgluaIaMetGlyValPheAlaCysArgValPro 101
         |||||::| ||||| ||| ||||| ||| |||||
Db      401 CTGCATGACGCTCAGGACACTTGTATCTGGCCACACTGTTGCTCACCCTCCAGTCCCA 460
         |||||::| ||||| ||| ||||| ||| |||||
Oy      102 GlySantYrYrSerSerPheAspValGluLeuHIScYsLySValaSnIlePleu 121
         |||||::| ||| ::||| ||| ||| ::| |||
Db      461 GGACTCTACCAAGTTGGATTTCACATTTGAAGCTGTCCACAGGCTGTAAGTAGAGCTC 520
         |||||::| ||| ::||| ||| ||| ::| |||
Oy      122 MetArgLySglInleuAlaSnIySgluIuIleSerLySglInSerIleGInglu 141
         |||||::| ||| ::||| ||| ||| ::| |||
Db      521 ATGGAATGCGACCCCAAGTCATGAGAGAGAAAGCGAGGCCACGATGCTATGAGCAC 580
         |||||::| ||| ::||| ||| ||| ::| |||
Oy      142 ValThr 143
         ::|||
Db      581 ATTTCGA 586

RESULT 7
CHKCX 2609 bp DNA linear VRT 28-APR-1993
LOCUS CHKCX
DEFINITION Chicken type X collagen gene.
ACCESSION M13496
VERSION M13496.1 GI:211699
KEYWORDS collagen.
          Chicken red blood cell DNA, clone pYN92E1; and embryo chondrocyte,
          cDNA to mRNA, clone pYN3116.
          Gallus gallus
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
          Phasianinae; Gallus.
REFERENCE 1 (bases 1 to 2609)
AUTHORS Nimmla, Y., Gordon, M., van der Rest, M., Schmid, T., Linsemayer, T.
          and Olsen, B. R.
TITLE The developmentally regulated type X collagen gene contains a long
JOURNAL open reading frame without introns
MEDLINE J. Biol. Chem. 261 (11), 5041-5050 (1986)
PUBMED 86168227
COMMENT [1] has as yet not determined the start codon of the collagen X
          gene. The open reading frame starts at position 166. There are
          several 'atg' codons (starting at positions 190, 214, 256, 274 and
          310) following the 'taa' at position 163.
          Location/Qualifiers
          1..2609
          /organism="Gallus gallus"
          /db_xref="taxon:9031"
          /cell_type="chondrocyte"
          <1..379
          <380..2208
FEATURES
source

```


VERSION	AX332258.1	GI:18122892
KEYWORDS		
SOURCE	human.	
ORGANISM	Homo sapiens	
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. 1	
AUTHORS	Young,P.E., Augustus,M., Carter,K.C., Ebner,R., Endress,G., Horridge,S., Soppel,D.R. and Weaver,Z. Cancer gene determination and therapeutic screening using signature gene sets Patent: WO 0194629-A 2767 13-DEC-2001;	
JOURNAL	Avalon Pharmaceuticals (US) Location/Qualifiers 1..2235 /organism="Homo sapiens" /db_xref="taxon:9606"	
FEATURES		
source		
BASE COUNT	524 a 632 c 693 g 386 t	
ORIGIN		
Alignment Scores:		
Pred. No.:	1.87e-05	2235
Score:	163.50	Matches: 42
Percent Similarity:	42.14%	Conservative: 17
Best Local Similarity:	30.00%	Mismatches: 48
Query Match:	20.70%	Indels: 33
DB:	6	Gaps: 5
US-09-997-610-2_COPY_1_149 (1-149) x AX332258 (1-2235)		
QY	19 GlyProProlAlaHisProArgProProGluIValGlyProProGlyAlaProGlyLeu	38
Db	1660 GGCCAGCGTGGCTTCACAGACCCTCCAGGCCCTCCAGACCTCCAGAGCCCCCAGCTGTG	1719
QY	38 -----	38
Db	1720 ATGCCCCCTTAACACCACCCACGGAGAGATATGTCGCACATATGGGGCTGGGAATTGAT	1779
QY	39 -----ProIntfThrgLysIuLleserIgumethrLysCysProCyspro	54
Db	1780 GGGCGTAAACCCCACATGCTACGGGG-----GCTAAGAAGGCAAGATGAGGGCCA	1833
QY	55 AsPIlEGluArSerAlaPherThrValLysIeuSerIgtLysLeuProIeuProPhelys	74
Db	1834 GCCATGAGTAGTGCCTCATTTACCCGACGCTAACCGCACCTTTCCACCGGTGGGGGCG	1893
QY	75 ProIleIlePherThrGlyValIeuTrpAsnAlaGlnArgaspLeuylsgLIAlaMetgly	94
Db	1894 CCAATGAGATTTAACAACACTGCTGTATACGGCAGACAGAACTACAACCCGACAGACGG	1953
QY	95 ValPheAlaCysArgValProGlyAsnTyrtyrSerSerPheaspValGIuLeuHisHis	114
Db	1954 ATCTTACACCTGTAGAGTGCTCTGTGTCTACTACTTTGCTATACACGTT-----CAC	2004
QY	115 CysLys---ValasnlIetRpleuMeCarLySGInIleuAlaasnLysGIuIuIle	133
Db	2005 TGCAGGGGGGGGAGCACTGTGGGTT-----GCTCATATCAAGACACAGAGCCCGTGG	2055
RESULT 10		
LOCUS	HSCOLBA1	2235 bp mRNA linear PRI 04-AUG-1992
DEFINITION	Human COLBA1 mRNA for alpha I(VIII) collagen.	
ACCESSION	X57527	
VERSION	X57527.1	GI:30081
KEYWORDS	collagen alpha I type III; extracellular matrix protein.	
SOURCE	Homo sapiens.	
ORGANISM	Homo sapiens	
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. 1 (bases 1 to 2235)	
AUTHORS	Olsen,B.R.	
TITLE	Direct Submissions	
JOURNAL	Submitted (05-FEB-1991) B.R.Olsen, Dept of Anatomy and, Cellular	

REFERENCE	2	(bases 1 to 2235)
AUTHORS	Muregaki,Y., Mattei,M.G., Yamaguchi,N., Olsen,B.R. and Nilmola,Y.	
TITLE	The complete primary structure of the human alpha 1 (VIII) chain and assignment of its gene (COL8A1) to chromosome 3	
JOURNAL	Eur. J. Biochem. 197 (3), 615-622 (1991)	
MEDLINE	91231001	
PUBMED	2029894	
FEATURES		
source	Location/Qualifiers	
gene	1..2235	
CDS	/organism="Homo sapiens"	
	/db_xref="taxon:9606"	
	/chromosome="3"	
	/map="3q12-3q13.1"	
	/clone="Ym801"	
	/clone_11b="genomic in EMBL"	
	1..2235	
	/gene="COL8A1"	
	1..2235	
	/gene="COL8A1"	
	/codon_start=1	
	/product="alpha 1(VIII) collagen"	
	/protein_id="CAA0748.1"	
	/db_xref="GI:30082"	
	/db_xref="SWISS-PROT:P27658"	
	/translation="MAVLPGPLQLGLVLTLSLSIRLLOAGVYGIKPLPPIQPM PQIKQYQPLQGVPHMLPLAKDGLAMKEMPHLOAYKEYPHLPYMKELTOPAPMKRKE AVRKQYELPLASLRGDEGPERGPRGPRGLGPHGIPGIKPKPQGVYKGRK CKMGKPRKPGKMGKAGKGLGQKGEIGPMKIPGPGPRGPRGLGIGKPKGGLPGQ PEPKDPRKPLPGQGLRGKDKGDKGFGMGAPGVKGPFGMLGLPVLGLPGVKKPY TGFPGDPLGKLPAPAPPEPKQGLTGVPGVQSPGPIGLGKPGQGLTGGPGFPGKKG EGGLGLGLGAPGLGLGIGKPGFPGKGRMGVAPALGPRGKPGIISPGIGSGRGE GLPGLPGMPGPAIGFPGKPGEGGIYVPGQPGPGKGLGQFPKGLGLVYVPGPG KMGFPGPIGKGEHGGKGVPLGLVPELGLGKKEPGLPDGGLGPGPIGLIGSGSGR TGPPIIPPKKEPGLPGPRPGTIGKRGVAGLHGPRGKRGALPGQGGGLGPRGPRG PPGPAVNPPTPPQGETLPDMGLGIDGVAEPHATGAKKGNKGFATEMPFLMELTA PPPPVGGPVKFNKLINGRQNYNPQGTITFCIEVPGYYTAYHHCKGNVWVAFKNN BEVMTYDEYKKKFLDQASGAVLLLRPGDVRVFLQMPSSQAAGLYAGQVYVSFGYVL LYPM"	
sig_peptide	1..84	
	/gene="COL8A1"	
mat_peptide	85..2232	
	/gene="COL8A1"	
	/product="alpha 1(VIII) collagen"	
BASE COUNT	524 a 632 c 693 g 386 t	
ORIGIN		
Alignment Scores:		
Pred. No.:	1.87e-05	Length: 2235
Score:	163.50	Matches: 42
Percent Similarity:	42.14%	Conservative: 17
Best local Similarity:	30.00%	Mismatches: 48
Query Match:	20.70%	Indels: 33
DB:	9	Gaps: 5
US-09-997-610-2_COPY_1_149 (1-149) x HSCOL8A1 (1-2235)		
QY 19	GLYProProLahIsProArProProGluLulValGlyProProGlyAlaProGlyLeu	38
Db 1660	GGCGAGCTGGCCCTTCACGAGACCCCAAGCCCTCCAGACCTCCAGACCCCAAGCTGTG	1719
QY 38	-----	38
Db 1720	ATGCCCCCTACACACCACCCAGGAGAGATATGCCAGATATGGGGCTGGGAATTGAT	1779
QY 39	-----ProGlnTyrThrGlyGluLeuSerGluMetThrIysCysProCysPro	54
Db 1780	GGCGTGAACCCCTTCATCTACGGGG-----GCTAAGAAAGCAAGANTGAGAGGCCA	1833
QY 55	AspIleGluArgSerAlaPheThrValIysLeuSerGlyIysLeuProLeuProPheIys	74
Db 1834	GCCATAGAGATGGCTCATTTACCGCCAGAGCTAACCGCAACCTTTCCACCGGTGGGGGC	1893

Oy	75	ProIIIEPheThglValleuYrksnslaglnAqASpleuylGluAlametyl	94
Dd	1894	CCAGCAGATTACAAACTGCTGTATACGGCACACAGAACTCAACC	CACAGAGCC 1953
Oy	95	ValPheaIcysArgValproGLysAntTyTyrSerSerpheAspValGluleuhis	114
Dd	1954	ATCTTCACCTCGTAGCTCCCTGTCTACTACTTTGCATTCACACAGTT-----CAC	2004
Oy	115	CysLys---ValAsnIIETripLeumetArGylsgInlIeLeuAlAsnLysGluIule	133
Dd	2005	TGCAGGGGGGAGACGTGTGGTT-----GCTCATTCACAGACAAGACAGCCCTG	2055
RESULT 11			
RABCOLVIII			
LOCUS	RABCOLVIII	2415 bp	mRNA linear MAM 27-Apr-1993
DEFINITION	Rabbit type VIII collagen alpha-1 chain mRNA, complete cds.		
ACCESSION	J05042		
VERSION	J05042.1 GI:164895		
KEYWORDS	alpha-1 type VIII collagen; basement membrane protein.		
SOURCE	Rabbit (strain New Zealand white) young adult cornea endothelial cells. cDNA to mRNA, clone pCE1240 and MK1.		
ORGANISM	Oryctolagus cuniculus		
REFERENCE	Eunariota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.		
AUTHORS	Yamauchi, N., Benya, P.D., van der Rest, M., and Ninomiya, Y.		
TITLE	The cloning and sequencing of alpha I(VIII) collagen cDNAs demonstrate that type VIII collagen is a short chain collagen and contains triple-helical and carboxyl-terminal non-triple-helical domains similar to those of type X collagen		
JOURNAL	J. Biol. Chem. 264 (27), 16022-16029		(1989)
MEDLINE	89380199		
PUBMED	2476437		
COMMENT	Draft entry and computer-readable copy of sequence [1] kindly submitted by Y.Ninomiya 23-AUG-89.		
FEATURES			
source	location/Qualifiers		
	1..2415		
	/organism="Oryctolagus cuniculus"		
	/db_xref="taxon:9986"		
CDS	181..2415		
	/note="alpha-1 (VIII) collagen precursor"		
	/codon_start=1		
	/protein_id="AAA31204.1"		
	/db_xref="GI:164896"		
	/translation="MAVPRGPOLQLVLITISGSITFLIAGAATYGKIPRPPQTPOM PPQIPQIQPLAQOYPNHPRLAKDGLTMKEKNHAYGKEIFHLIDPYMKLEVPVRMGE AVPKRGKETPLASLRGEGRPGERPGRRPRLPDQIGIKGKGDPGVGRKE GMPPMRGPMGAMPGRGAKAGEIGRGSELGIRMPDRPGRICTGKRGSGRLPQ PGAAGDRGPKRPRGPILOGRKSEKSGMGSLRLKRGKNNRPSRGVGKAPRGPG TGFGPQGRLKRGPRGEPGPOPIGVAGOGSRGLGYVKRGDGIIRDGFRGCKG EGLGILGPRPGILRGKRGPRGPKRGKDRGIGVNGALGPRGSEKRVARMGSPRRP GLDETRPMGRPDAIGFPRGKRGESBGIVLPDGPRGPKERPELDGPRGKFIPLVEGPRG IRGLDPTIRKGEAGNHKLRLGVLPRGRLILPKRGKGIIRSDGLDGPGLRGLGPRG IGRPGIIRGKGEPLRGPRGPRGPRGKGVAGLIRPGRCALPQSQGLRGPGRPG PCGPRAPMPRTPAPOGELIPDMGLGIDGIVTAHYAAKKKNGSRVAEMAFDELTA PEWTVGAPIREDLLYNGRONYNQGITFLICEVGVYVRAVHNHGKGGNVVALFKNN EPWWTYTDEYKKKGLDQASGSAVILLRGRDVFLOMSEDAAGLYAQQYVHSSTGL LYPM"		
	181..240		
	/note="alpha-1 (VIII) collagen signal peptide"		
	241..2412		
	/product="alpha-1 (VIII) collagen"		
BASE COUNT	536 a 719 c 392 t		
ORIGIN			
Alignment Scores:			
Pred. No.:	2.03e-05	Length:	2415
Score:	163.50	Matches:	45
Percent Similarity:	44.20%	Mismatched:	16
Best Local Similarity:	32.61%	Gaps:	29
Query Match:	20.70%	Gaps:	7

```

US-09-997-610-2_COPY1_1149 (1-149) x RABOLV1111 (1-2415)

Oy 19 GTPProPAlaHisProArGrProBProGlUGluValAlGlyProProGlyAlaProGlyLeu 38
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1840 GGCACGACCTGGCCCTCCAGGCCCCCAGGCCCTCCAGGCCCTCCAGGCCCTCCAGGCCCTG 1899
Oy 39 -----ProGlnTyrThrGlyGlu---HisSerGlnMet----- 48
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1900 ATGCCCCGACACCGACACCCAGGAGAGATGTGCGAGATATGAGGGCTGGGATTGAT 1959
Oy 49 ---ThrLysCysPro-----CysProAspIle 56
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1960 GGGGTAATAAACCTCCATCCAGCCCTATGCGGCCAAGAAAGCAACAGCGGGCGCAGCCTAT 2019
Oy 57 GUAArgSerAlaPheThrValLysLeuSerGlyLysLeuProLeuProPheLysProIle 76
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2020 GAGAGCGCTGACCTTACCGCCGACGTGACGACACCCCTTCCCGCGGTGGGGCGCCGATA 2079
Oy 77 IlePheThrGlyValLeuTyrAsnIleGlnA9AspLeuLysGluAlaMetGlyValPhe 96
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2080 AAGTCGACAGACCTCTCTACATGCGACAGCAACTACACCCGACAGCGGCATCTTC 2139
Oy 97 AlaCArgValProGlyAsnTyrTyrSerSerPheAspValGluLeuHisLysCysIlys 116
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2140 ACGTCGAGAGTTCCCGGGGTCTACTACTTGGCTACACAGCT-----CACGTGCAAG 2190
Oy 117 ---ValAsnIleTyrLeuMetArgLysGlnIleLeuAlaAsnLysGluGluIle 133
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2191 GGGGCGAACGCTGTGGGT-----GCTGTGTCAGAGAACACAGACGCGCGTG 2235

RESULT 12
BC013581 2506 bp mRNA linear PRI 07-SEP-2001
LOCUS BC013581 Homo sapiens, clone MGC:9568 IMAGE:3875911, mRNA, complete cds.
DEFINITION BC013581
ACCESSION BC013581.1 GI:15488903
VERSION MGC.
KEYWORDS Homo sapiens.
SOURCE Homo sapiens.
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
1 (bases 1 to 2506)
Strausberg, R.
Direct Submission
Submitted (04-SEP-2001) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
NIH-MGC Project URL: http://mgc.nci.nih.gov
Contact: MGC help desk
Email: gcnbs-remail.nih.gov
Tissue Procurement: DCTD/DP/GenZdar
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LINL)
DNA Sequencing by: Sequencing group at the Stanford Human Genome
Center, Stanford University School of Medicine, Stanford, CA 94305
Web Site: http://www.shgc.stanford.edu
Contact: (Dickson, Mark) mcdexall@stanford.edu
Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers,
R. M.
Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LINL at: http://image.llnl.gov
Series: IRAK Plate: 14 Row: a Column: 8
This clone was selected for full length sequencing because it
passed the following selection criteria: matched mRNA gi: 996814.
Location/Qualifiers
1..2506
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="MGC:9568 IMAGE:3875911"
/tissue_type="Lung, large cell carcinoma"

```

```

CDS
    /clone_1lb="NIH_MGC_68"
    /lab_host="DH10B"
    /note="Vector: pCMV-Sport6"
    238..2472
    /codon_start=1
    /product="Unknown (protein for MGC:9568)"
    /db_xref="GI:15488904"
    /protein_id="AAH13581.1"

BASE COUNT      624 a      698 c      753 g      431 t
ORIGIN
Alignment Scores:
Pred. No.:      2.6e-05      Length:      2506
Score:          162.50      Matches:      42
Percent Similarity: 42.75%      Conservative: 17
Best Local Similarity: 30.43%      Mismatches: 50
Query Match:      20.57%      Indels:      29
DB:              9      Gaps:      5

US-09-997-610-2_copy_1_149 (1-149) x BC013581 (1-2506)
QY      19  G|P|P|P|P|A|H|S|P|R|A|G|P|R|P|P|G|L|V|A|G|L|P|R|P|G|L|A|P|R|G|L|Y|L|E|U      38
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      1897 GGCACGACCTGGCGCTTCCAGGACCCCGAGCCCTCCAGGACCTCCAGGACCCCGACGCTGTG      1956
QY      39  -----ProGlnTyrThrGlyGluIle-----
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      1957 ATGCCCCCTACACCCACACCCCGAGAGATCTGCCAGATATGGGGCTGGGAATTGAT      2016
QY      46  -----SergLumethrLysCysProCysProAspIle      56
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      2017 GCGGTGAACCCCCCATACCTACGGGGCTAAGCAAGAGATGAGAGCGCATCTAT      2076
QY      57  G|U|A|G|S|E|R|A|P|H|E|T|H|V|A|L|Y|S|E|U|S|E|R|G|L|Y|S|E|U|P|R|O|L|E|U|R|O|P|H|E|L|Y|S|P|R|O|L|E      76
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      2077 GAGATGCGTGCATTTACCGCGGAGCTAACGCGACCTTTCCACCGGTGGGGGCCCGAGTG      2136
QY      77  I|E|P|H|E|R|H|G|L|Y|V|A|L|E|U|R|Y|A|S|N|A|G|L|A|R|A|G|S|P|L|E|U|S|G|L|U|A|L|E|U|G|L|Y|A|L|P|H|E      96
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      2137 AAGTTTAAACAACCTGCTATTAACGACAGACACACCAACCCGACGACAGCATCTTC      2196
QY      97  A|A|C|Y|S|A|G|V|A|P|R|G|L|Y|A|N|T|Y|R|T|Y|S|E|R|S|E|R|P|H|E|S|P|A|I|G|L|E|U|H|S|I|S|C|Y|S|Y|S      116
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      2197 ACCTGTGAGTGCCCTGCTGTACTACTTTCATACACACCTT-----CACTCAAG      2247
QY      117 ---ValanIletrPleuMetArgLysGlnIleLeuAlasnlYsgLnuGluIle      133
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      2248 GGGGGCAACGTGTGGTT-----GCTCTATTCAAGAACAGACGACCGGTG      2292

RESULT 13
AC069222/c
LOCUS      AC069222      117000 bp      DNA      linear      PRI 29-MAR-2002
DEFINITION Homo sapiens 3 BAC RP11-383123 (Roswell Park Cancer Institute Human
ACCESSION AC069222
VERSION    AC069222.23      GI:19033388
KEYWORDS   HTG.
SOURCE     Homo sapiens.
ORGANISM   Homo sapiens
```

```

REFERENCE
AUTHORS
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 117000)
Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
Alsbrooks,S.L., Amarantunge,H.C., Are,J.R., Ayele,M., Banks,T.,
Barbata,J., Benton,J., Blmage,K., Blankenburg,K., Bonini,D.,
Bouck,J., Bowls,S., Brieve,M., Brown,E., Brown,M., Bryant,N.P.,
Burch,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
Davila,M.L., Davis,C., Davy-Carroll,L., Dedrich,D.A.,
Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
Douthwaite,K.J., Diaper,H., Dugar-Kocha,S., Durbin,K.J.,
Earhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
Gabisi,A., Gao,J., Garcia,A., Garner,T., Garra,N., Gill,R.,
Gorelli,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
Harris,C., Harris,K., Hart,M., Havlak,P., Hayes,A., He,X.,
Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C.,
Hollins,B., Homs,F., Howard,S., Huber,J., Huliyk,S., Hume,J.,
Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,
Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,
Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,
Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
Louiseged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,
Martinez,E., Massey,E., Mawhney,E., Mcleod,M.P., Meador,M.,
Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabate,K.,
Moore,S., Morgan,M., Moorish,T., Morris,S., Moser,M., Neal,D.,
Nelson,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N., Nguyen,N.,
Nickerson,E., Nwokenkwo,S., Ogun,M., Okunodu,G., Otagunye,N.,
Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L., Peters,L.,
Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y., Rives,M.,
Rojas,A., Rojibokan,I., Rolfe,M., Rulz,S., Savery,G., Scherer,S.,
Scott,G., Shen,H., Shooshtrani,N., Stison,I., Sodergren,E.,
Sonaike,T., Sparks,A., Stanley,H., Stone,H., Sutton,A., Statak,A.,
Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H., Tansey,J., Taylor,C.,
Taylor,T., Telford,B., Thomas,N., Thomas,S., Usmani,K., Vasquez,L.,
Vera,V., Villalón,D., Vinson,R., Wang,O., Wang,S., Ward-Moore,S.,
Warren,R., Washington,C., Watlington,S., Williams,G.,
Williamson,A., Wiczysk,R., Wooden,S., Worley,K., Wu,C., Wu,Y.,
Wu,Y.F., Zhou,J., Zorilla,S., Naylor,S.L., Weinstein,G. and
Gibbs,R.
Direct Submission
Unpublished
2 (bases 1 to 117000)
Worley,K.C.
Direct Submission
Submitted (22-MAY-2000) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 117000)
Worley,K.C.
Direct Submission
Submitted (28-FEB-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
4 (bases 1 to 117000)
Worley,K.C.
Direct Submission
Submitted (01-MAR-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
5 (bases 1 to 117000)
Worley,K.C.
Direct Submission
Submitted (29-MAR-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Mar 1 2002 this sequence version replaced gi:18958589.
INFORMATION: http://www.ngsc.bcm.tmc.edu/ or email
gc-help@bcm.tmc.edu
```



```

repeat_region 13699..13810
/note="MIR repeat: matches 76..174 of consensus"
repeat_region 13806..13919
/note="MIR repeat: matches 77..189 of consensus"
repeat_region 13945..14060
/note="MIR repeat: matches 24..142 of consensus"
repeat_region 14061..14367
/note="AluY repeat: matches 1..301 of consensus"
repeat_region 14368..14452
/note="MIR repeat: matches 141..225 of consensus"
repeat_region 14589..14679
/note="MIR repeat: matches 173..262 of consensus"
misc_feature 14597..15201
/note="match: GSS: Em:A053482"
misc_feature 14616..15060
/note="match: GSS: Em:A0370601"
repeat_region 14868..15040
/note="MIR repeat: matches 49..233 of consensus"
repeat_region 15071..15188
/note="L2 repeat: matches 2112..2239 of consensus"
repeat_region 15304..15399
/note="MLTIB repeat: matches 1..99 of consensus"
repeat_region 15490..15662
/note="AluSg1 repeat: matches 2..114 of consensus"
repeat_region 15669..15727
/note="MLTIB repeat: matches 119..178 of consensus"
repeat_region 15728..16027
/note="AluSc repeat: matches 1..299 of consensus"
repeat_region 16028..16245
/note="MLTIB repeat: matches 178..390 of consensus"
repeat_region 16546..16854
/note="AluY repeat: matches 1..300 of consensus"
repeat_region 18296..18323
/note="MSTR repeat: matches 2..29 of consensus"
repeat_region 18324..18392
/note="MER66-internal repeat: matches 4919..4993 of consensus"
repeat_region 18393..18712
/note="AluDb repeat: matches 1..311 of consensus"
repeat_region 18713..19133
/note="MER66-internal repeat: matches 4548..4919 of consensus"
misc_feature complement(18872..19230)
/note="match: GSS: Em:A005063"
misc_feature 19251..19719
/note="match: GSS: Em:BI4179"
repeat_region 19537..20290
/note="HERVH21 repeat: matches 4657..5784 of consensus"
repeat_region 20317..20382
/note="33 copies 2 mer ta 68 conserved"
repeat_region 20513..20666
/note="77 copies 2 mer tt 70 conserved"
repeat_region 20682..21008
/note="AluSg1 repeat: matches 1..306 of consensus"
repeat_region 21239..21553
/note="HUKS-P3 repeat: matches 4410..4713 of consensus"
repeat_region 21882..22254
/note="THB1B repeat: matches 1..364 of consensus"
repeat_region 22302..22537
/note="MER66-internal repeat: matches 2186..2417 of consensus"
repeat_region 22538..22850
/note="AluSp repeat: matches 1..313 of consensus"
repeat_region 22851..23801
/note="MER66-internal repeat: matches 1210..2186 of consensus"
repeat_region 23905..23989
/note="MER66-internal repeat: matches 3017..3102 of

```

Alignment Scores:
Pred. No.:
Score:
Percent Similarity:

0.00201 Length: 145880
162.50 Matches: 42
50.89% Conservative: 15

```

Best Local Similarity: 37.50% Mismatches: 46
Query Match: 20.57% Indels: 9
DB: 9 Gaps: 3
US-09-997-610-2_COPY_1_149 (1-149) x HS302D9 (1-145880)
QY 37 G1yleuProGlnTyrThrGlyGluIleSerGluMetThrLysCysProGlyProAspIle 56
||||| :||| :||| :||| :||| :|||
DB 90953 GGAATCCCGTGTTCATGCAAA-----CAAAGCCAGAGCAGTACAGAGCATATT 90900
QY 57 Glu-----ArgSerAlaPheThrValLysLeuSerGlyLysLeuPro 70
||| ||||| :||| :||| :||| :|||
DB 90899 GAAAGAGTTCATCCCAACCAAGATCGATTGTGCTGAATCATGATCGACCCCA 90840
QY 71 LeuProPheLysProIleIlePheThrGlyValLeuTyrAsnAlaGlnArgSpleuLys 90
:||||| :||||| :||| :||| :||| :|||
DB 90839 GTGCCCTTCTACGCCCTTCTCTTCAAGAGAGCCCTTATTAATACATGATTCATTCAT 90780
QY 91 GluAlaMetGlyValPheAlaCysArgValProGlyAsnTyrTyrSerSerPheAspVal 110
||| ||| :||| :||| :||| :|||
DB 90779 TTCTCTGCGTAGAGATTCACCTGTACATCCCTAGTGTACCATCTTGTTGATATT 90720
QY 111 GluLeuH1SH1ScysLysValAsnIleTyrPleuMetArgLysGlnIle--LeuAlaAsn 129
||||| :||| :||| :||| :||| :|||
DB 90719 GAGCTGTTTCAGAGCTGTGCAATATGGGTATTATGAGCAATGATGATGAGAAAT 90660
QY 130 LysGluGluIleSerLysGlnIleSerIleGlnGlu 141
||| :||| :||| :||| :|||
DB 90659 AAACATGCTGTGCGCAAGATGCTATGACGAGAA 90624
RESULT 15
AC120598 152354 bp DNA linear HTG 23-JUL-2002
LOCUS Homo sapiens, *** SEQUENCING IN PROGRESS ***, 53 unordered pieces.
AC120598
ACCESSION AC120598.3 GI:21902634
VERSION HTG: HTGS.PHASE1.
KEYWORDS
SOURCE Homo sapiens.
ORANISM Homo sapiens.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 152354)
REFERENCE
AUTHORS Albrooks,S.L., Amaralunge,H.C., Are,J.R., Ayete,M., Banks,T.,
Barbarta,J., Benton,J., Blmage,K., Blankenburg,K., Bonnin,D.,
Bouck,J., Bowle,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
Bunay,C., Burch,P., Burrell,K.L., Byrd,N.C.,
Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
Devila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
DeJaney,K.R., Delgado,O., Dem,A.L., Ding,Y., Dinn,H.H.,
Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escoto,M.,
Falls,T., Ferraguto,D., Flagg,N., Ford,R., Foster,P., Frantz,P.,
Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
Correll,J.H., Guevara,M., Gunaratne,P., Hate,S., Hamilton,K.,
Harris,C., Harris,K., Hart,M., Havila,P., Hawes,A., Hernandez,J.,
Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,
Homsi,F., Howard,S., Huber,J., Huliy,S., Hume,J., Jackson,L.E.,
Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudh,S.,
Karlsone,E., Kelly,S., Khan,U., King,L., Korvab,J., Kovar,C.,
Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Louised,H.,
Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
Maheshwari,M., Mapua,P., Martin,R., Martinale,A., Martinez,E.,
Massey,E., Mashiney,E., McLeod,M.P., Meador,M., Mei,C., Metzger,M.,
Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N.,
Nguyen,N., Nickerson,E., Nwokenkwo,S., Ogun,M., Okunodu,G.,
Oraunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
Peters,L., Pickens,R., Primus,E., Pu,L.L., Qulies,M., Ren,Y.,
Rives,M., Rojas,A., Rojupokan,I., Rolfe,M., Rutz,S., Savery,G.,

```


THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: February 19, 2003, 14:53:22 ; Search time 3433.83 Seconds

(without alignments)
2084.671 Million cell updates/sec

Title: US-09-997-610-2_COPY_18_459
Perfect score: 2416
Sequence: 1 AGPPAHPRPPEVGPACG.....GQFPFGTAFTQCLYLHGM 442

Scoring table:
BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 16154066 segs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=x1h
-O=/cgn2_1/USPTO.spool/US09997610/runat_10022003_160823_23843/app-query.fasta.1.1635
-DB=EST -QFWT=fastap -SUFFIX=1st -MINMATCH=0.1 -LOOPEL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blomsu62 -TRANS=human40.cdi -LIST=45
-DOCLIGN=200 -THR=score-pct -THR.MAX=100 -THR.MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pic -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09997610.ecgn.1.1.181 -runat_10022003_160823_23843 -NCPU=6 -ICPU=3
-NO_XLPHY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMOUT=120
-WARN_TIMOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -Fgapop=6 -Fgapext=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

EST:*
1: em_estba:*
2: em_esthum:*
3: em_estlin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hlc:*
9: gb_est1:*
10: gb_est2:*
11: gb_hlc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vit:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rtd:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result		Query		ALIGNMENTS		Description	
No.	Score	Match	Length	DB	ID		
1	1184.5	49.0	2615	10	BE420422	BE420422 32-393 hu	
2	1070	44.3	2009	11	BC004496	BC004496 Homo sapi	
3	1012.5	41.9	1514	10	BE512633	BE512633 32-1514 h	
4	902.5	37.4	898	13	BM457166	BM457166 AGENCOURT	
5	877	36.3	937	14	BO893011	BO893011 AGENCOURT	
6	869.5	36.0	1036	13	BM471183	BM471183 AGENCOURT	
7	838	34.7	1050	13	BM472108	BM472108 AGENCOURT	
8	821	34.0	876	14	BQ423563	BQ423563 AGENCOURT	
9	802.5	33.2	870	13	B1488505	B1488505 603021014	
10	793	32.8	941	14	BQ723415	BQ723415 AGENCOURT	
11	752.5	31.1	1076	13	BM554723	BM554723 AGENCOURT	
12	735	30.4	1080	14	BQ422247	BQ422247 AGENCOURT	
13	733	30.3	699	9	AL602279	AL602279 DKFZP686B	
14	732	30.3	900	13	BM445699	BM445699 AGENCOURT	
15	731	30.3	998	14	BM903785	BM903785 AGENCOURT	
16	718.5	29.7	697	17	AG018926	AG018926 Homo sapi	
17	718	29.7	736	13	B1754555	B1754555 603023384	
18	716.5	29.7	721	12	BE728616	BE728616 601563827	
19	715	29.6	654	17	AG112666	AG112666 Pan trogl	
20	714.5	29.6	760	12	BG281182	BG281182 602402107	
21	709	29.3	717	17	AG086951	AG086951 Pan trogl	
22	700.5	29.0	704	12	AG020199	AG020199 CIT-HSP-2	
23	693.5	28.7	940	12	BG679573	BG679573 602627616	
24	690.5	28.6	955	12	BC681769	BC681769 602627877	
25	688.5	28.5	937	14	BQ223504	BQ223504 AGENCOURT	
26	681	28.2	704	17	AG058828	AG058828 CITR1-E1	
27	680.5	28.2	1063	14	BM926985	BM926985 AGENCOURT	
28	678.5	28.1	1021	17	AG0374295	AG0374295 RPCI11-14	
29	676	28.0	701	17	AG058873	AG058873 CITR1-E1	
30	673	27.9	885	12	BC761979	BC761979 602718930	
31	670	27.7	833	9	AU140358	AU140358 AU140358	
32	664	27.5	676	17	AG064424	AG064424 Pan trogl	
33	664	27.5	858	9	AU122114	AU122114 AU122114	
34	662	27.4	691	17	AG062447	AG062447 Pan trogl	
35	660	27.3	720	17	AG116938	AG116938 Pan trogl	
36	658.5	27.3	681	17	AG116668	AG116668 Pan trogl	
37	656	27.2	729	13	BM045873	BM045873 603624875	
38	653	27.0	622	12	BG108072	BG108072 602280006	
39	651	26.9	772	9	AU117241	AU117241 AU117241	
40	647	26.8	677	17	AG097258	AG097258 Pan trogl	
41	642.5	26.6	739	17	AG001404	AG001404 Homo sapi	
42	635.5	26.3	596	10	BE390947	BE390947 601283117	
43	630	26.1	602	17	AG0374537	AG0374537 RPCI11-1	
44	627.5	26.0	623	17	AG075663	AG075663 Pan trogl	
45	626	25.9	639	17	AG115932	AG115932 Pan trogl	

RESULT 1
BE420422
LOCUS BE420422
DEFINITION BE420422 2615 bp mRNA linear EST 11-OCT-2001
32-393 human bone marrow cDNA library Homo sapiens cDNA, mRNA
sequence.
ACCESSION BE420422
VERSION BE420422.1 GI:16041640
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 2615)
AUTHORS Lu, X., Cui, L., and Li, Y.
TITLE DDRT-PCR from B cell

JOURNAL

Unpublished (2000)
 Contact: xingwu Lu, lianxian Cui, yonghai Li
 Department of Biochemistry
 Institute of Basic Medical Science, Peking Union Medical College
 Dongdan Sanliao 5, Beijing, P.R.C, 100005
 Tel: 86-010-65296951
 Email: luxingwu@263.net
 full-length and coding sequence.

FEATURES

SOURCE

Location/Qualifiers

1. 2615
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="human bone marrow CDNA library"
 /tissue_type="bone marrow"
 /note="Organ: uterus; Vector: pOTB7; Site 1: XhoI; Site 2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCAAGAG(6). Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies)."

BASE COUNT 705 a 531 c 641 g 738 t
 ORIGIN

Alignment Scores:

Pred. No.: 1.23e-96 Length: 2615
 Score: 1184.50 Matches: 244
 Percent Similarity: 74.65% Conservative: 21
 Best Local Similarity: 68.73% Mismatches: 76
 Query Match: 49.03% Indels: 15
 DB: 10 Gaps: 3

US-09-997-610-2_COPY_18_459 (1-442) x BE420422 (1-2615)

QY 101 AsnIleTrrpleuMetAryGlySGInIleLeuAlaAsnLysGIuGIuIleSerLysGIuIn 120
 DB 494 AACCTTGAGCTTGAGAGATCATTTAGGGTACTGGCAGAAATAATTTCTAAGCAGCA 553
 QY 121 SerIleGInIuValThrTrpValLeuLeuLysAlaPheSerPheIleArgIuIaGlu 140
 DB 554 AGCATTCAGAGGTRACTGGGTCTGTAAAGGCATTTGTTTAAAGGAGAGAGG 613
 QY 141 HsLysSerSerGIuAsnLeuHsProAspAsnValIleLysLysLysAsnProPheSer 160
 DB 614 CATGAAGTTGGAAATAATTTGCACCTGCACATGACATAGAAAAGAAATCCCATATCT 673
 QY 161 GIuGIuLysPheLysLeuAlaAlaGIuIleCysIleCysAsnIuIuLeuAsnValAsn 180
 DB 674 GAGGAGAAATCAAGCTGGCTGCAGAAATTTGCATATGTAACAAGAGATTAAATGTTAAT 733
 QY 181 ProGlnAspAsnGIuAsnIleSerTrpThrcysGlnArserSerGlnGlnSerIle 200
 DB 734 CCCCAAGCAATGGGAAATATGTTCCAGGAGCATGCCAGAGGTCTTACACGACGCCCTC 793
 QY 201 LysSerLeuAlaTrpArgProArgArgLysTrp---PheCysGIuThrcylProGIlySer 219
 DB 794 CCATCAGAGGCTGCAGGCTCAGAGGAAAAAGATCTCTGGGCCAGGCCAGGGGTCC 853
 QY 220 LeuGIuLysValGlnProArgAspLeuValProCysValProValAsnSerAla----- 237
 DB 854 CCATCTGTGTGTCAGCTGGGGACTGTGGCTTGTCTCCACCTGCTCCAGCCATGCT 913
 QY 238 -----ValAlaSerGIuGIuLysLeuLysLeuProLysProLysProLys 248
 DB 914 GAAAGGGCCAAATAGAGCTCGGATTTGGCTTCAGAGGGGTCCAAAGCTGTAAAGCTTGG 973
 QY 249 GlnLeuProSerGIuValGIuProValGIuAlaLysLysSerArgIleGIuValTrpGIu 268
 DB 974 CAGCTTCATGTTGTTGAGCTGCAGAGTGCACAGAAATGAGGTTGGGAA 1033
 QY 269 ProProlleArgPheGlnLysIleTyrGIuAsnProTrpMetProArgGlnLysPheAla 288
 DB 1034 CCTTGCTAGATTTCAGAAAGATGTATGGAAATGCCGTGATGCCGCGCAGAGTCTGCT 1093

QY 289 ValGIuValGIuSerSerTrpArgThrSerAlaArgValAlaGlnLysGIuAsnValGIu 308
 DB 1094 GTAGGGAGGGGCCCTCATGAGATCTCTCTAGGGCAGTACAGAAAGGAAATGTGGGA 1153
 QY 309 TrpGIuProProHsArgValProSerGIuAlaProSerSerArgAlaValArgArgSer 328
 DB 1154 TCGAGACCCCAACAGATCCCTACTGGGGCAGCCGCTGTGGAGTGTGAGAAAGAGG 1213
 QY 329 ProProSerSerArgLeuGIuLysGIuArgSerThrAspSerLeuGlnHsValProGIu 348
 DB 1214 CCACCATCTCCAGACCCCAAGATGATTCACATGACGTTTGCTCCATGCACCTGGA 1273
 QY 349 LysSerThrAspThrcylCysGlnProValLysAlaAlaLysMetGlnSerValProTyr 368
 DB 1274 AAACCGCAGAC---AACAC-CAGCCCATGAACACACTAGGAGGAGGCTGTACCTGC 1329
 QY 369 LysThrValValAlaGIuLeuThrLysThrValGIuLyleTyrLeuLeuHsCysHsAsp 388
 DB 1330 AAAGCCACAGGGGGAGCTGCCCAAGATCATGGGAACCCACTTGTGATCATCAGATGAC 1389
 QY 389 LeuAspValArgHsGIuValLysArgAspHsPheGIuAlaLeuArgPheAspCysPro 408
 DB 1390 CTGGATGTAGATTGTGGACTTAAAGGAGATCATTTTGAGCTTTAAAGATTGACTGCC 1449
 QY 409 ThrGIuPheArgThrTyrMetGIuProVal-ProLeuCysPheGIuGlnPheProPh 428
 DB 1450 TCTAGATTTCAGACTTCATGGAGGGCTGTAGGCCCTTGTGTTACTCATATTCTCCATT 1509
 QY 428 eGIuThrAlaValPheThrGlnCysLeuTyrLeuHsCysMet 442
 DB 1510 TGGACAGCTGTGTTTACACATACCTGTACCCCATGTGATTC 1552

RESULT 2

BC004496 BC004496 2009 bp mRNA hmc 12-JUL-2001
 LOCUS Homo sapiens, similar to hypothetical protein FLJ14058, clone
 DEFINITION IMAGE:383113, mRNA.

ACCESSION BC004496
 VERSION BC004496.1 GI:14709139

KEYWORDS htc.
 SOURCE Homo sapiens.

ORGANISM Homo sapiens.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 2009)
 TITLE Direct Submission

AUTHORS Strausberg,R.
 JOURNAL Submitted (12-MAR-2001) National Institutes of Health, Mammalian
 Gene Collection (MGC), Cancer Genomics Office, National Cancer
 Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
 USA

REMARK NIH-MGC Project URL: <http://mgc.ncl.nih.gov>
 COMMENT Contact: MGC help desk
 Email: cgabbs-remail.nih.gov

Tissue Procurement: ATCC/DCTD/DRP
 CDNA Library Preparation: Rubin Laboratory

CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Institute for Systems Biology

<http://www.systemsbio.org>
 contact: amadan@systemsbiology.org

Anup Madan, Rachel Dickhoff, Jessica Fahy, Stephanie Ford, Julia
 Greene, Mark Kettelman and Anuradha Madan

Clone distribution: MGC clone distribution information can be found
 through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
 Series: IRML Plate: 14 Row: d Column: 7
 This clone was selected for full length sequencing because it
 passed the following selection criteria: Hexamer frequency ORF
 analysis

This clone has the following problem: frame shifted.

FEATURES Location/Qualifiers

source I. 2009

[illegible]

DB	Accession	Source	Organism	Reference	Authors	Title	Journal	Comment
Db	1210	GGCGGAGACATTCACACCA	GGCCCGTGGAAAGACCA	GGGAGGAGATATATCTGCANA	1269			
Oy	370	thralvalalaglauldeuthrlysthrValaglyletrylaenleuHiscysHiaspLeu	389					
Db	1270	GGCAGAGGAGAGAGCTTCCCAAGACTAGTGGAGACCTACTCTTACATCAGACGACCTT	1329					
Oy	390	AspValaIrgHISGLYValIylsArGAspHISpHeGlyAlaLeuArGpHeAspCysProthr	409					
Db	1330	TGTGTGACCGCTCGAGTCAAGAGACTATTTTGGAGCTTTAAATCTCAGCCCTGCT	1389					
Oy	410	GlyPheArgThrTyrMetIlyProVal -ProLeuGlyPheGlyGlnPhePheProPheI	429					
Db	1390	GGATTTCAACATTCGCAATGGGCGCTACACACCCCTTTGTTTGGCCGATTTCTCCAAATTGG	1449					
Oy	429	YThrAlaValPheThrGlnCysIleuTyrLeuHiscysMet	442					
Db	1450	AAAGCGCTGATTTTACCACCAATACCTGTACCCCATTTGTATC	1489					
RESULT 3	BE512633	1514 bp	mRNA	linear	EST 11-OCT-2001			
LOCUS	BE512633	32-1514	human B lymphocyte cDNA library	Homo sapiens	CDNA, mRNA			
DEFINITION	BE512633	sequence.						
ACCESSION	BE512633	GI:16041645						
VERSION	EST.							
KEYWORDS	human.							
SOURCE	human.							
ORGANISM	Homo sapiens							
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;							
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.							
TITLE	1 (bases 1 to 1514)							
JOURNAL	Lu, X., Cui, L., and Li, Y.							
COMMENT	DDRT-PCR from B cell							
	Unpublished (2000)							
	Contact: xingwu lu, liangxian Cui, yonghai Li							
	Department of Biochemistry							
	Institute of Basic Medical Science, Peking Union Medical College							
	Dongdan Santiao 5, Beijing, P.R.C, 100005							
	Tel: 86-010-65296951							
	Email: luxingwu@263.net							
	Full cDNA sequence.							
FEATURES	Location/Qualifiers							
source	1..1514							
	/organism="Homo sapiens"							
	/db_xref="taxon:9606"							
	/clone_lib="human B lymphocyte cDNA library"							
	/tissue_type="bone marrow"							
	/note="Organ: Tongue; Vector: pAMP10; mRNA made from							
	tongue epithelium, cDNA made by oligo-dT priming.							
	Non-directionally cloned into uDG sites. Size-selected on							
	agarose gel, average insert size 500 bp. Primary library.							
	CDNA Library Preparation: David B. Krizman, Ph.D.							
	REFERENCE: Krizman et al. (1996) Cancer Research							
	56:5380-5383."							
BASE COUNT	380 a 341 c 398 g 395 t							
ORIGIN								
Alignment Scores:								
Pred. No.:	2,396-81	Length:	1514					
Score:	1012.50	Matches:	207					
Percent Similarity:	74.26%	Conservative:	18					
Best local Similarity:	68.32%	Mismatches:	64					
Query Match:	41.91%	Indels:	15					
DB:	10	Gaps:	3					
US-09-997-610-2-COPY_18_459 (1-442) x BE512633 (1-1514)								
Oy	153	IleLysLysLysAspProPheSerGlnGlyLysPheLysLeuAlaGluIleCysIle	172					
Db	1	ATAGAAAGAAATCCCATATTTCTGTGAGGAGAAATCAACCTGCTGCAGAAATTTGCATA	60					
Oy	173	CysAsnGluGluLeuAsnValAsn						

Db	61	TTGTAACAGAGAGTAAATGTTATATCCCAACACATGCGGAAATGCTCCAGGCGATC	120
Qy	193	GLNArgSerSerGlnGlnSerIleLysSerLeuAlaTrpArgProArgIyStrp---	211
Db	121	CAGAGGCTTTCACAGCAGCGCCCTCCACATCACAGGCGCTGACGGCTTAGAGAAAGTGA	180
Qy	212	PheGlySerThrGlnProGlySerIleuGlySerGlyValGlnProArgAspLeuValProGly	231
Db	181	TTCTCTTGCGGACGACGCCAGGATGCCCATGCTGTGTGCGCTGGGACATGGTGCCCTGT	240
Qy	232	ValProValAsnSerAla-----ValAlaSer	240
Db	241	GTCCACAGCTGCTCCAGCCATGGCTGAAGGGGCCAACATGAGCTCGAATTGGCTTCA	300
Qy	241	GLuGlyAlaSerProLysProTrpGlnLeuProSerGlyValGluProValGlyAlaLys	260
Db	301	GAGGCTGCAAGCTGTAGAGCTTTGGAGCTTGCATGCTGTGTGGCTGCGACTGCACAG	360
Qy	261	LysSerArgGlnLeuValTrpGluProIleArgPheGlnLysIleTyrGlyAsnPro	280
Db	361	AAGTCAAAATTTGGGTTTGGAGACCTGCGCTGATTTGGAAGATGTGGAATGGC	420
Qy	281	TrpMetProArgGlnLysPheAlaValGlyAlaGlySerSerTrpArgTrpSerAlaArg	300
Db	421	TGCAATGCCCGGACGAGACTCTGCTTAGGGGAGGGGCCCTCATGAGAAATCTCTCTAG	480
Qy	301	ValValGlnLysGlnValGlnValGlyTrpGluProProHisArgValProSerGlyAlaPro	320
Db	481	GCAATACAGAGGGAATGTGGATCGAGGCCCAAGCAGAAATCCCTACTGGGCGACCG	540
Qy	321	SerSerArgAlaValArgArgSerProProSerSerArgLeuGlnLysGlyArgSerThr	340
Db	541	CGTAGTGAAGCTGTGAGAAAGGCGCACCATCTCCACAGCCCGAATGTATGATCACT	600
Qy	341	AspSerLeuGlnHisValProGlnLysSerThrAspTrpGlnGlyGlnProValLysAla	360
Db	601	GACGAGCTTGTCTACCATCCCTGGAAAACCGCAGC--AAAC-CAGGCCATGAAAGA	656
Qy	361	AlaGlyMetGlnSerValProTyrLysThrValAlaValAlaGlnLeuThrLysThrValGly	380
Db	657	GCTAGGCAAGAGCTGTACCTCGCAAAAGCACAGGGGGGAGCTGCCAAGATCATGGGA	716
Qy	381	IleTyrLeuLeuHisGlySerHisAspLeuAspValArgHisGlyValLysValArgAspHisPhe	400
Db	717	ACCACACCTCTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATTTT	776
Qy	401	GlyValAlaLeuArgPheAspCysProThrGlnLysPheArgTrpTrpMetGlyProVal-Prole	420
Db	777	GGAAGCTTAAAGATTTTACTGCCCCCTCTGATATTTCAGACTTCAGATGGGGGCTGTAGCCCT	836
Qy	420	uCyAspGlnGlyGlnPhePheProPheGlnLysAlaValPheThrGlnCysLeuTyrLeuHis	440
Db	837	TTGTTTACTCAATTTCTCCCATTTGGAAACAGCTGTGTTTACACAAATACCTGATCCCA	896
Qy	440	sCyMet 442	
Db	897	TTGTATC 903	
RESULT 4			
LOCUS	BM457166	898 bp	MRNA linear EST 05-FEB-2002
DEFINITION	AGENCOURT_6411690 NIH_MGC_92	Homo sapiens	cdNA clone IMAGE:5583427
ACCESSION	5', mRNA sequence.		
VERSION	BM457166		
KEYWORDS	BM457166.1 GI:18506206		
SOURCE	EST.		
ORGANISM	human.		
	Homo sapiens		
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
	Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.		
REFERENCE	1 (bases 1 to 898)		
AUTHORS	NIH-MGC http://mgc.ncl.nih.gov/ .		

TITLE	National Institutes of Health, Mammalian Gene Collection (MGC)			
JOURNAL	Unpublished (1999)			
COMMENT	Contact: Robert Strausberg, Ph.D. Email: cgapbs-remail.nih.gov Tissue Procurement: ATCC CDNA Library Preparation: Life Technologies, Inc. CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Agencourt Bioscience Corporation Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov Plate: LLAM12346 row: h column: 20 High quality sequence start: 9 High quality sequence stop: 713. Location/Qualifiers			
FEATURES	1..898 /organism="Homo sapiens" /db_xref="taxon:9606" /clone="IMAGE:5583427" /clone.lib="NIH_MGC_92" /tissue.type="embryonal carcinoma, cell line" /lab.host="DH10B (phage-resistant)" /note="Organ: testis; Vector: pCMV-Sport6; Site.1: NotI Site.2: SalI; Cloned unidirectionally; oligo-dT primed. Average insert size 2.5 kb. Library enriched for full-length clones and constructed by Life Technologies Note: this is a NIH_MGC Library."			
BASE COUNT	225 a	228 c	254 g	191 t
ORIGIN				
Alignment Scores:				
Pred. No.:	1.le-71	Length:	898	
Score:	902.50	Matches:	188	
Percent Similarity:	70.83%	Conservative:	16	
Best Local Similarity:	65.28%	Mismatches:	64	
Query Match:	37.36%	Indels:	21	
DB:	13	gaps:	3	
US-09-997-610-2_COPY18_459 (1-442) x BM457166 (1-898)				
QY	171	CysIleCysAsnIuIuLeuAsnValAsnProGlnAspAsnIyGluAsnIleSerTrp	190	
		:::::		
DB	35	TGCTATTAGCAGCAGGAGGCTTATTTATCCACATGCGGAAATGCTCCAGG	94	
QY	191	ThrcysGlnArgSerSergInuSerIleIySerIeuAlaTrpArgProArgArg---	209	
	:::::		:::	
DB	95	TCATGTCAGAGACCTTCATGATGAGCCCTCCATCACAGGCGCCAGGCGCCAGGAGAA	154	
QY	210	LysTrpPheGlySerGlyhncIyProGlySerIeuCysValGlnProArgAspLeuVal	229	
DB	155	AAGGTGTTTGTGTGGGTGGGCCCAAGGTCGTGTGTGTGTGACGCTTAGGGACTTGGTG	214	
QY	230	ProCysValProValAsnSer-----AlaVal	238	
DB	215	CCCTGTGTCCACGCTCTCCACGATGGCTGAAGAGGCCACATAGAGCTTGGGCTGAG	274	
QY	239	AlaSerGluGlyAlaSerProIySerProTrpGlnIleuProSerGlyValGluProValGly	258	
DB	275	GCTTCACAGAGGATGAGGCCCAAGGATTTGGAGCTTCCATGTAGTATGAGCCTGTGGGT	334	
QY	259	AlaIySyrSerArgIleGluValTrpIuProProIleArgPheGlnIySrlIyGly	278	
DB	335	GCAAGATGTCACGAATTTAGATTTGGGAGACCTCTGCGCTTACGAATATATATGA	394	
QY	279	AsnProTrpMetProArgGlnIySyrPheAlaValGlyValGlySerSerTrpArgThrSer	298	
DB	395	AACACCTCGGATGTCACGCAAAAGTTTCTCCAGGCGGACAGAGCCCTCACAAGGGAACTCT	454	
QY	299	AlaArgValValIlnIySyrGlyAsnValGlyTrpIuProProIleArgValProSerGly	318	
DB	455	CCTAGGGCAGGTGGGAGGGAATGTGGGGCA-CAGGCCCCACACAGAGATCCCTGTGGG	513	
QY	319	AlaProSerSerArgAlaValArgArgSerProProSer-----	331	


```
Db 514 GCACGTGCTAGTGGAGCTGTGGAGAGGCCACGATGCCAGAACGGTAGACCCAGAT 573
Qy 332 SerArgLeuGlnLysGlyArgSerThrAspSerLeuGlnHisValProGluLysSerThr 351
Db 574 GGTGACACCCAGAAATGGTAGATCAACACCACTTGATCTTGTCCTGGAAACCTTACA 633
Qy 352 AspThrGlnCysGlnProValLysAlaIaGlyMetGlnSerValProTyrLysThrVal 371
Db 634 GACACTCAACATCCACTGGAGAAACAGCCAGAGGAGGAGCTGTACTCTGCCAAACCCACA 693
Qy 372 ValAlaGluLeuThrLysThrValGlyIleTyrLeuLeuHisCysHisAspLeuAspVal 391
Db 694 GGGGGGAGAGCTGCACAAAGACCATGGAACTATCTTCATCAGCATCAGACCTGAATGTG 753
Qy 392 ArgHisGlyValLysArgAspHisPheGlyAlaLeuArgPheAspCysProThrGlyPhe 411
Db 754 AGACATGGAGTCMAAGGAGATCATATGTAGACTTTMAAATTTGACTGCCCCACTGATTTT 813
Qy 412 ArgThrTyrMetGlyProVal-ProLeuCysPheGlyGlnPhePheProPheGlyThrAl 431
Db 814 CAGACTTCATAGACCTCGTAACCCCTTGTTGTGGCAATTTCTCTCTTTGGAAACAGC 873
Qy 431 aValPheThrGlnCysLeuTyr 438
Db 874 TGTATTTACCAATACCTGTAC 895

RESULT 5
B0893011
LOCUS B0893011 937 bp mRNA linear EST 16-AUG-2002
DEFINITION AGNCOURT: 8122304 Lupski dorsal root ganglion Homo sapiens cDNA
clone IMAGE:6179271 5', mRNA sequence.
ACCESSION B0893011
VERSION B0893011.1 GI:22285025
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 937)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLES National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: rgs@bbs-remail.nih.gov
Tissue Procurement: Dr. James R. Lupski
cDNA Library Preparation: Life Technologies, Inc.
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/MLML at:
http://image.llnl.gov
Plate: LHAM13560 row: C column: 16
High quality sequence stop: 722.
Location/Qualifiers
1. 937
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:6179271"
/clone_lib="Lupski_dorsal_root_ganglion"
/sex="male"
/tissue_type="dorsal root ganglia"
/dev_stage="adult, 36 yr"
/lab_host="DH10B"
/note="Vector: pCMV-SPORT6 (Life Technologies); Site_1:
NotI; Site_2: SalI; cDNA made by oligo-dT priming.
Directionally cloned using the following adaptors:
5'-TCGACCGACGGGCGG-3' and
5'-GACTAGTTCTTAGATCGGACGCGCCCTT(15)-3'. Size selected >
1 kb for average insert length 1.7 kb. This is a primary
library, non-amplified. Library constructed by Life
Technologies and donated by J. Lupski, M.D./Ph.D. (Baylor
College of Medicine) and is available through Life
```

```
BASE COUNT 235 a 226 c 258 g 214 t 4 others
ORIGIN
Alignment Scores:
Pred. No.: 2,46e-69 Length: 937
Score: 877.00 Matches: 194
Percent Similarity: 70.53% Conservative: 19
Best Local Similarity: 64.24% Mismatches: 64
Query Match: 36.30% Indels: 28
DB: 14 Gaps: 4
US-09-997-610-2_COPY_18_459 (1-442) x B0893011 (1-937)
Qy 156 LysAsnProPheSerGlnGlyLysPheLysLeuAlaIaGlnIleCysAlleCysAsnGlu 175
Db 6 AAAATTTGATTTTTCGAATAGAAATTCAAGCCAGCTGCAGAAATTTGATTAAGGAAGG 65
Qy 176 GluLeuAsnValAsnProGlnAspAsnGlyGluAsnIleSerTrpThrCysGlnArgSer 195
Db 66 GAGCCAAATGTTTAATCCCAAGACAAATGGGAAATGTCTCCAGGCCATGTAGAGGCT 125
Qy 196 SerGlnIleSerIleLysSerLeuAlaTrpArgProArgArg--LysTrpPheCysGly 214
Db 126 TCACGCGAGTCCCTCCATCAGAGGCCAGAAAGCTAGAGGAAAGAGGTTTCATGCGG 185
Qy 215 ThrGlyProGlySerLeuCysCysValGlnProArgAspLeuValProCysValProVal 234
Db 186 CTGGGCCAGAGGCTCCCAAGCATATGTGAGCTGG-GGTTTGCTCAGCTGATCCAGCT 244
Qy 235 AsnSerAlaValAlaSer-----GluGlyAla 243
Db 245 GCCCCAGCTGTGCTGAAAGGCGACAAATGTAGACCTCGGCCCTTGCTTACATGTGCA 304
Qy 244 SerProLysProTrp-----GlnLeuProSerGlyValGluProValGlyAlaLysLys 261
Db 305 AGCCCCATGCTGTGTGGTGTAGCTTCATGTGGTGTGACCTCGAGGTGCACAGAAAG 364
Qy 262 SerArgIleGluValITrGluProProIleArgPheGlnLysIleTyrGlyAsnProTrp 281
Db 365 TCMAAATTTGAGCTTTTGAACTCCACCTCAATTTTCAGAAATGTAGAAATCCCTGG 424
Qy 282 MetProArgGlnLysPheAlaValAlaGlyLysSerSerTrpArgThrSerAlaArgVal 301
Db 425 ATGCCCAGGCAAAATTTTGCTGCAAGGCTGGGCTCTCATGAGAACTCTCTGAGGCA 484
Qy 302 ValGlnLysGlyAsnValGlyTrpGluProProHisArgValProSerGlyAlaProSer 321
Db 485 GTGCAGAAAGGAAATGTGGGGTGGAGTCCCAAGAGTCCCTACATGGGGCA----- 538
Qy 322 SerArgAlaValArgArgSerProProSerSerArgLeuGlnLysGlyArgSerThrAsp 341
Db 539 -----CTGTCTCCCAAGACC-CAGAAAGTAGAATCCACTAGC 573
Qy 342 SerLeuGlnHisValProGluLysSerThrAspTrpGlnCysGlnProValLysAlaIa 361
Db 574 ACCTTGACACTGTGTGCTGGAAAGCCACAGACACTCAATGGCCACTTAGAAACAAAC 633
Qy 362 GlyMetGlnSerValProTyrLysThrValAlaIaGluLeuThrLysThrValGlyIle 381
Db 634 AGGAGGAGAGACTAGCCCTGCCAAAGCCACAGGCGCAGAGCTCTGACAGCACTGGGAAC 693
Qy 382 TyrLeuLeuHisCysHisAspLeuAspValArgHisGlyValLysArgAspHisPheLys 401
Db 694 TACCTCTGTATCAGACATGACCTGATGTGAGACATGAGAGTGAAGGAAACATTTTGA 753
Qy 402 AlaLeuArgPheAspCysProThrGlyPheArgTrpTyrMetGlyProVal-ProLeu 421
Db 754 GCTTAAGATTTGAGCTGCGCTGCTGATTTGAGCTGCAATGAGGCCCTGTAGCCCTTTG 813
Qy 421 sPheGlyGlnPhePheProPheGlyThrAlaValPheThrGlnCysLeuTyrLeuHisCys 441
Db 814 ATTGGCCAGATTTCSSCATTTTGAATGGATATTTTAC-CATGCTCATGACSSCATTTG 872
```

OY	441	smeC_4442
	1:::	
D6	873	TATC_876
RESULT 6		
LOCUS	BM471183	
DEFINITION	BM471183	1036 bp mRNA linear EST 05-FEB-2002
ACCESSION	AGENCOURT_6478314 NIH_MGC_72	Homo sapiens cDNA clone IMAGE:5563049
VERSION	5,	mRNA sequence.
KEYWORDS	BM471183.1	GI:18520225
SOURCE	EST.	
ORGANISM	human.	
REFERENCE	Homo sapiens	
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
TITLE	Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.	
JOURNAL	1 (bases 1 to 1036)	
COMMENT	NIH-MGC http://mgs.nci.nih.gov/. National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished (1999)	
	Contact: Robert Strausberg, Ph.D. Email: cgabbs@email.nhl.gov CDNA Library Preparation: Life Technologies, Inc. CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Agencourt Bioscience Corporation Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov Plate: LLAM12293 row: g column: 18 High quality sequence stop: 661.	
FEATURES	Location/Qualifiers	
Source	1..1036	
	/organism="Homo sapiens"	
	/db_xref="taxon:9606"	
	/clone="IMAGE:5563049"	
	/clone_lib="NIH_MGC_72"	
	/tissue_type="melanolic melanoma"	
	/lab_host="DH10B (phage-resistant)"	
	/note="Organ: skin; Vector: PCMV-SPORT6; Site:1: NotI; site:2: SalI; Cloned unidirectionally. Primer: Oligo dT. Average insert size 2 kb. Library constructed by Life Technologies."	
BASE COUNT	254 a 239 c 294 g 247 t	2 others
ORIGIN		
Alignment Scores:		
Pred. No.:	1.38e-68	Length: 1036
Score:	869.50	Matches: 177
Percent Similarity:	73.0%	Conservative: 24
Best Local Similarity:	64.368	Mismatches: 60
Query Match:	35.9%	Indels: 15
DB:	13	Gaps: 3
US-09-997-610-2_COPY_18_459 (1-442) x BM471183 (1-1036)		
OY	178	AsnValAsnProGlnAspAsnGlyAlaSerTrpPheCysGlnArgSerGln 197
		:::
D6	33	AATTATATCCCAACAACCATGGGGAATAATGTCTCCACTTCCTGTCAGAGACCTTCACGTG 92
		:::
OY	198	GlnSerIleLeuSerLeuAlaTrpArgProArgArg--LysTrpPheCysGlyThrGly 216
		:::
D6	93	CAGACCCTCCCATCACAGGCCCAAGAGGCCCAAGAGAAAAAGTGTTCTGTGGCGTGGGT 152
		:::
OY	217	ProGlySerLeuCysCysValGlnProArgAspLeuValProCysValPro----- 233
		:::
D6	153	TCAAGGTCGCCCATGCTGTGTGAAGCCTTAGGAGCACTTGAGCCCTGTGCCAAGCTGCTCCA 212
		:::
OY	234	-----ValAsnSerAlaValAlaSerGluGlyAlaSerPro 245
		:::
D6	213	GCCTTGCTGAAGGGGCTAATGTATACCTCTGGCTGGCTTGAAGGCTGAAGAGCCCC 272

QY	246	LysrProtrpbln---	LeuProserGtlyValGluProValAGlyAlaLysLysSerArgIle	264
Db	273	AAGCTTTGGAGGCTTTTCCATGCTGCTGTGGCCCTGTGGTGGACAGATCAAGAAACT	332	
QY	265	GlUValrTgbluProProIleArGheGlnLysrIleTryGlyAsnProTfPmeProArg	284	
Db	333	GAGGTTTGGAACTCT-GCTAGATTCCAGAAAGATGTTAGMAATGCCGTGGATGCCAGG	391	
QY	285	GlnLysPheAlaValAGlyAlaGlySerSerTrpArgThrSeraIaArgValaIaGlnLys	304	
Db	392	CAAAAGTTTCTCCAGAGGTGGGGCCCTCAGTAGAGAAACCTCTGCTAGGGACAGTGTGGAA	451	
QY	305	GlyAsnValAGlyTrpLupProPhisArgValProSerGlyAlaProSerSerArgAla	324	
Db	452	GGAATGTGGGTGGAGCCCGACACAGTGGCCCTGTGGGGGACACGCTTACGTGGAGCT	511	
QY	325	ValArgTrGserProProSerSerArgLeuGlnLysGlyArgSerThrAspSerLeuGln	344	
Db	512	GTGAGGAAGAGGGCCACCATCTCTCCAGACCCCAAGATGTTAGATCATCAACACAGCTTGCA	571	
QY	345	HisValProGlnLysSerThrAspThrGlnCysGlnProValLysAlaAlaGlyMetGln	364	
Db	572	TGTGCACCTTGGAAAGCCGACACACTGCAGACCCAGCCCATGAAAGCAGCTGGAGAGAG	631	
QY	365	SerValProTryLysThrValaIaAGluLeuThrLysThrValAGlyIleTryLeuLeu	384	
Db	632	GCTTACCCCTGCAAAACACAGAGGGCAGAGCGCTTAAAGACATGGAGAACCCACTCTTG	691	
QY	385	HisGlyHisAspLeuAspValArgHisAGlyValLysAspHisPheGlyAlaLeuArg	404	
Db	692	CATGACGATCACTGGATATGAGACTGGAGCAAGAGATCATTTTGGAGCTTTGAA	751	
QY	405	PheAspCysProThrLysPheArgThrTryMetGlyProVal-ProLeuCysPheGlyI	424	
Db	752	TTTGATTTGCCCGCCGTGGATTTCAAGACTGTGTGANNCCCTGATGAACCCCTTTGTTGGCA	811	
QY	424	nPhePheProPheGlyThrAlaValPheThrGlnCysLeuTry	438	
Db	812	ATTCTCCATTTTGAATGCTACATTACCAATACCTGTAC	854	
RESULT 7				
LOCUS	BM472108			
DEFINITION	BM472108	1050 bp	mRNA	linear
ACCESSION	AGENCOURT_6465359	NTIH_MGC_72	Hom	sapiens
VERSION	BM472108			
KEYWORDS	BM472108.1	GI:18521150		
SOURCE	EST.			
ORGANISM	human.			
REFERENCE	Homo sapiens			
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
TITLE	Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.			
JOURNAL	National Institutes of Health, Mammalian Gene Collection (MGC)			
COMMENT	Unpublished (1999)			
	Contact: Robert Strausberg, Ph.D.			
	Email: cgrabbs@mail.nih.gov			
	Tissue Procurement: ATCC/DCTD/DP			
	CDNA Library Preparation: Life Technologies, Inc.			
	CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)			
	DNA Sequencing by: Agencourt Bioscience Corporation			
	Clone distribution: MGC clone distribution information can be			
	found through the I.M.A.G.E. Consortium/LLNL at:			
	http://image.llnl.gov			
	Plate: LHAM12233	row: m	column: 14	
	High quality sequence stop: 658.			
FEATURES				
source				
	1..1050			
	Location/Qualifiers			
	/organism="Homo sapiens"			
	/db_xref="taxon:9606"			
	/clone_image="5359381"			
	/clone_id="NIH_MGC_72"			

/tissue_type="melanotic melanoma"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: skin; Vector: pCMV-SPORT6; Site_1: NotI;
 Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
 Average insert size 2 kb. Library constructed by Life
 Technologies."

BASE COUNT 295 a 262 c 272 g 221 t
 ORIGIN

Alignment Scores:

Pred. No.: 1.03e-65 Length: 1050
 Score: 838.00 Matches: 197
 Percent Similarity: 62.22% Conservative: 27
 Best Local Similarity: 54.72% Mismatches: 79
 Query Match: 34.69% Indels: 57
 DB: 13 Gaps: 9

US-09-997-610-2_COPY_18_459 (1-442) x BM472108 (1-1050)

Qy 114 GUGUUISESerLysGInGInSerIleGInGInUValThrTPValIleuLeuLysAlaPhe 133
 |||||
 Db 17 GAAGAAATTTTATGACAGCAAGCATTTCAAGATTGACTTGGTCTCTTAAAGGCATTC 76
 Qy 134 SerPheIleArgIuAlaGInHisLysSerSerGluAsnLeuHisProAspAsnValIle 153
 |||||
 Db 77 AGATTATTAAAGGAAGCAGACGATAAAGTTACAGAAATTTGACAGCTGACAAATGATATA 136
 Qy 154 LysLysLysAsnProPheSerSerGluGlyLysPheLysLeuAlaIleGlyIleCysIleCys 173
 ::::|
 Db 137 GAGAGAGAAAGCCCATTTTGGAGAGAAATTTCAAGCCGATGATTAATTTGGCTAAGT 196
 Qy 174 AsnGluGluLeuAsnValAsnProGInAspAsnGlyGluAsnIleSerTrpThcCysGln 193
 |||||
 Db 197 AACAGAAAGCCGATGTTAATCCCAACAGACGAGAGACATCTCCAGACATATCAG 256
 Qy 194 ArgSerSerGInGInSerIleLysSerLeuAlaTrpArgProArgArgLysTrpPheCys 213
 |||||
 Db 257 AGGCTTCACAGACGCCCTCCGTCACAGAGCCAGAGGCTAGAGAAATGGTTT--- 313
 Qy 214 GlyThrGlyProGlySerLeuGlyCysValGInProArgAspLeuValProCysValPro 233
 |||||
 Db 314 -----CCTGGACCT-----GTGCAGCCTAGTACTTGGCACCCTGTCTCCCA 355
 Qy 234 ValAsnSer-----AlaValAlaSerGluGly 242
 |||||
 Db 356 GCTACTCCAGTAGTGGCTGAAGAGGGCCAAAGTACAGCTTACGCCATGGCTTCAGAGGCT 415
 Qy 243 AlaSerProLysProTrpGInLeuProSerGlyValGInProValGlyAlaLysLysSer 262
 |||||
 Db 416 GCAAGCGCCAGACCTTGCGACCTCCGATGCTGTGAGCTGCMAATGCACAGACGTCGA 475
 Qy 263 ArgIleGluValTrpGluProProIleArgPheGInLysIleTyrGlyAsnProTrpMet 282
 |||||
 Db 476 AGATTGGATTGGGAACCTCTGCTAGATTTCAGAGATGACGGAAGTCCCGGATG 535
 Qy 283 ProArgGInLysPheAlaValGlyValGlySerSerTrpArgTrpSerAlaArgValVal 302
 |||||
 Db 536 CCCAGGCAAAAGTTGCTGCGAGGGGCGAGGCGCTTCATGAGAAACCTCTGCTAGGCGAATG 595
 Qy 303 GInLysGInLysValGlyTrpGluProProHisArgValProSerGlyAlaProSerSer 322
 |||||
 Db 596 TGGAGAGGAACATGGGGTGTGAGCCCCCAGACAGAAATCCCTACTGAGGACACACACTAGT 655
 Qy 323 ArgAlaValArgArgSerProProSerSerArgLeuGInLysGlyArgSerTrpAspSer 342
 |||||
 Db 656 GGACTCTGAGAGAAAGGCCCATCTCTCCACACCCCAAGATGATGATTCACCAAGCAGC 715
 Qy 343 LeuGInHisValProGluLysSerThrAspThrGInCysGInProValAlaAla--- 361
 |||||
 Db 716 TTGCACCATGACCCCTGGAAAGCCGACAGACATTCACACCCAGCCGGGGAAGCAGCCAAA 775
 Qy 362 -----GlyMetGInSerValProTyrLysThrVal 371
 |||||

Db 776 AGGGGGAATATATCTCCAGACCCAGAGGGGAGGAGCTCC----- 817
 Qy 372 ValAlaGluLeuThrLysThrValGlyIle-TyrIleuLeuHisCysHisAsp-LeuAsp 391
 |||||
 Db 818 -----CCCAACAGTATGGGAACCCCTACCTTACATTCGACAGACATGGGT 865
 Qy 391 aLArgHisGlyValLysArgAspHisPheGly-----AlaLeu 404
 ::::|
 Db 866 GAAACCTGGAAGTCCAAAGGGAACCTTTTGGAGAGCTTAAATCCTGACTGGCCCTG- 924
 Qy 404 rGpHeAspCysProThrGlyPheArgThrTyrMetGlyProVal-----ProLeuCysP 422
 |||||
 Db 925 -----CTGGATTTCCTCAAAATTCGATGGGCGCCCTTAAGACCCCTTGGGT 970
 Qy 422 heGlyGInPhePheProheGlyThrAla---Val-PheThrGInCysLeuTyr 438
 |||||
 Db 971 TGGGCCGATTCCTCCCAATTTGGAAAGGCGTGGTTTTCACCAATACCTGTTC 1024

RESULT 8
 LOCUS BQ423563 876 bp mRNA linear EST_23-MAY-2002
 DEFINITION AGENCOURT_7845919 NIH_MGC_72 Homo sapiens cDNA clone IMAGE:6064274
 5', mRNA sequence.
 ACCESSION BQ423563
 VERSION BQ423563.1 GI:21118878
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 876)
 AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cga@bbs.femail.nih.gov
 Tissue Procurement: ATCC/DC/DP
 CDNA Library Preparation: Life Technologies, Inc.
 DNA Library Arrayed by: The I.M.A.G.E. Consortium (LMNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LMNL at:
 http://image.llnl.gov
 Plate: LRAM1338 row: 1 column: 03
 High quality sequence stop: 674.
 FEATURES
 source
 Location/Qualifiers
 1..876
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:6064274"
 /clone_lib="NIH_MGC_72"
 /tissue_type="melanotic melanoma"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: skin; Vector: pCMV-SPORT6; Site_1: NotI;
 Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
 Average insert size 2 kb. Library constructed by Life
 Technologies."

BASE COUNT 242 a 200 c 246 g 187 t 1 others
 ORIGIN

Alignment Scores:
 Pred. No.: 2.76e-64 Length: 876
 Score: 821.00 Matches: 180
 Percent Similarity: 70.71% Conservative: 18
 Best Local Similarity: 64.29% Mismatches: 62
 Query Match: 33.98% Indels: 20
 DB: 14 Gaps: 4

US-09-997-610-2_COPY_18_459 (1-442) x BQ423563 (1-876)

Qy 98 CysLysValAsnIleTrp-----LeuMetArgLysGInIleLeuAlaAsnLysGlu 114
 |||||
 Db 43 TGCCCCAGAGATTGTGCACTTGTGAGAGATGATTTAGGTATCTGTGTAATA 102
 |||||

Db 544 CTGGGGCACTGCTAGTGAACCTGTGAGAGACACACCGCTCTCCAGACCCAGAAATG 603
OY 337 LYGASerThrAspSerLeuGlnHisValProGluLysSerThrAspThrGlnGlnP 357
Db 604 GTAAATCCACCAACAGCTGTGACCATGTGCTGAAAGATGACAGTATCATATCCAGC 663
OY 357 roValLysAlaAlaGluMetGluSerValProTyrLysThrValAlaGluLeuThrL 377
Db 664 CCATGAAGACGACCGAGAGAGGCTGTACCTGCAAAAGCCACAGGAGGAGCTGTCCA 723
OY 377 ystHValGlyIleTyrLeuLeuHisCysHisAsp-LeuAspValArgHisGlyValLys 396
Db 724 AGACCATGGGAACCA-CTCTTGATCAGCGTACCTGTGATGCGAGACCTGAGTCAAA 782
OY 397 ArgAspHisPheGlyAlaLeu-ArgPheAspCysProThrGlyPheArgThrTyrMeG 416
Db 783 GGAGATCATTTTGGAGCTTTAACAATGTGACCTGCTGCTGATTTGACACTTGCAATGG 842
OY 416 Y 416
Db 843 G 843
RESULT 10
B0723415 941 bp mRNA linear EST 16-JUL-2002
LOCUS AGENCOURT 8105155 Lupskl-sympathetic_trunk Homo sapiens cDNA clone
DEFINITION IMAGE:6193072 5', mRNA sequence.
ACCESSION B0723415 GI:21862312
VERSION B0723415.1
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 941)
AUTHORS NIH-MGC http://mgs.ncl.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: rgapds-remail.nih.gov
Tissue Procurement: Dr. James R. Lupskl
CDNA Library Preparation: Life Technologies, Inc.
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LMN)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/MLN at:
http://image.llnl.gov
Plate: LLMJ3596 row: b column: 17
High quality sequence stop: 606.
Location/Qualifiers
FEATURES
source
1. 941
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:6193072"
/clone_1ib="Lupskl-sympathetic_trunk"
/sex="male"
/tissue_type="sympathetic trunk"
/dev_stage="adult, 16 yr"
/lab_host="DH10B"
/note="Vector: pCMV-SPORT6 (Life Technologies); Site_1:
Note: Site_2: Salt; CDNA made by oligo-dT priming.
Directionally cloned using the following adaptors:
5'-TCGACCCAGCGGTCG-3' and
5'-GACTAGCTTCATGATCGGAGCGCGCCCT(15)-3'. Size selected >
1 kb for average insert length 1.9 kb. This is a primary
library, non-amplified. Library constructed by Life
Technologies and donated by J. Lupskl, M.D./Ph.D. (Baylor
College of Medicine); available through Life
Technologies."

Alignment Scores:
Pred. No.: 1.09e-61 Length: 941
Score: 793.00 Matches: 177
Percent Similarity: 68.51% Conservative: 21
Best Local Similarity: 61.25% Mismatches: 73
Query Match: 32.82% Indels: 18
DB: 14 Gaps: 2
US-09-997-610-2_copy_18_459 (1-442) x B0723415 (1-941)
OY 114 GUGLUILeserLysGlnGlnSerIleGlnGluValThrPvalLeuLeuLysAlaPhe 133
Db 51 GAAGAAATTTCTAACAGACAAAGATTCAAGAGCTCTCTGGCTGTTAAAGACATTG 110
OY 134 SerPheIleArgGluAlaGluHisLysSerSerGluAsnLeuHisProAspAsnValIle 153
Db 111 AGTTTTAAAGGAAACACAGCATAAAGTTGGAAATTTTCAGCCCTGACATGCAATA 170
OY 154 LysLysLysAsnProPheSerGlu-GlyLysPheLysLeuAlaGluIleCysIleCys 173
Db 171 GAAAGAAATATCCATTTTCTGAGGAGAAAGTCAAGCTGCTACAGAAATTTTACATAAG 230
OY 173 sAsnGluLysLeuAsnValAsnProGlnAspAsn-GlyGluAsnIleSerTyrPheCysG 193
Db 231 TAACAAGAAAGCCAAATGTAAATCCCAACAGATGGGGTAAATGTCTCAGAGCATATG 290
OY 193 InArgSerSerGlnGlnSerIleLysSerLeuAlaThrPargProArgarg---LysTrp 212
Db 291 AGAGGTCTTTCACGAGCCCTCCCTTTACAGCGCTGGGGGCTGAGAGAAAAAGTGGT 350
OY 212 heCysGlyThrGlyProGlnSerLeuCysCysValGlnProArgAspLeuValProCysP 232
Db 351 TTCGTGGGCTCAGACAGAGGCTCCCATGCTGTGTGAGCTTACAGGACTTGTGCTCGA 410
OY 232 aProval-----AsnSerAlaValAlaSerG 241
Db 411 TCACAAGCTGCTCAGCTGTGCTGAAGAGGGGTATCACAGACTCAGCTGTGGCTTAG 470
OY 241 LUGLYLaseProlYsProlTrrPglInLeuProSerGlyValGluProValGlyAlaLysL 261
Db 471 AGGGTCAAGTCCCAAGCCTTGACGCTTCCACGTGTGAGCTGCAAGTACATGCA 530
OY 261 ySerArgIleGluValTrrPglupProlIleArgPheGlnLysIleTrrGlyAsnProT 281
Db 531 ACTCAAGAAATTTGGGTTTGGAACTTCTGCTAGATTTTTCAGAGGTGCGAAATACCT 590
OY 281 rPmeProlArgGlnLysPheAlaValGlyLysSerTrrPargTrSerAlaArgV 301
Db 591 GGATGCCAGAGAGAAAGTTTGTTCAGAGGAGGAGGCGCTCATGAGAACTTCTGCTAG 650
OY 301 aLValGlnLysGlnAsnValGlyTrrPglupProlIleArgPheGlnLysIleTrrGly 320
Db 651 TATGTCGGAGGAGAAATGTAGGTGAGAGCCCAACAGATCCCTACTGAGACACTGC 710
OY 321 SerSerArgAlaValArgArgSerProProSerSerArgLeuGlnLys-GlyArgSerH 340
Db 711 CCTAGTGGAGCTGAGAGAAAGCCATCATCTTTCAGACCCAGAGAAAGTGTAGATCC 770
OY 340 rAspSer-LeuGlnHisValProGluLysSerThrAspThrGlnGlnPvalLysA 360
Db 771 TCACAGCTTGTGACGTGTGCTGGAAGCCACAGACACTCAAGCCAGCCATTAAG 830
OY 360 la-AlaGlyMetGluSerValProTyrLysThrValAlaGluLeuThrLysThrVal 379
Db 831 CACCCAGGAGGAGGAGCTGTACCTTCGCAAGACAGAGGGGTGGAGACCCCAAGACTAT 890
OY 380 GlyIleTyrLeuLeu 384
Db 891 GGAGCCACCTCTT 905
RESULT 11
BM554723 1076 bp mRNA linear EST 20-FEB-2002
LOCUS BM554723

DEFINITION AGENCOURT_6578521 NIH_MGC_41 Homo sapiens cDNA clone IMAGE:5468842
5', mRNA sequence.
ACCESSION BM554723
VERSION BM554723.1 GI:18794564
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 1076)
AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabs-r@mail.nih.gov
Tissue Procurement: DCTD/DRP
CDNA Library Preparation: Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: L1CM1973 row: b column: 11
High quality sequence stop: 665.
Location/Qualifiers

FEATURES
source
1. 1076
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_id="NIH_MGC_41"
/tissue_type="amelanotic melanoma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: skin; Vector: pORF7; Site:1: XhoI; Site:2:
EcoRI; cDNA made by oligo-dT priming. Directionally cloned
into EcoRI/XhoI sites using the following 5' adaptor:
GGCAGGAG(G). Library constructed by Ling Hong in the
laboratory of Gerald M. Rubin (University of California,
Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and
superscript II RT (Life Technologies). Note: this is a
NIH_MGC library."

BASE COUNT 274 a 251 c 295 g 256 t
ORIGIN

Alignment Scores:
Pred. No.: 6.44e-58 Length: 1076
Score: 752.50 Matches: 154
Percent Similarity: 68.75% Conservative: 11
Best Local Similarity: 64.17% Mismatches: 62
Query Match: 31.15% Indels: 13
Gaps: 1

US-09-997-610-2_copy_18_459 (1-442) x BM554723 (1-1076)

OY 216 GYPPROGLYSerLeucyGcysValGlnProARgaspLeuValProCyValProValasn 235
DB 5 GGCCCAAGGCTCCCTGCTGTGTACAGCCTAGAGGACTTGCTGCTGCCCAAGCCT 64
OY 236 SerAla-----ValAlaSerGlnGlnAlaSer 244
DB 65 CCAGCCATGGGTGAAGGACCAACATAGAGCTAGAGCTGTGGCTTCAGAGAGTCCAAAGC 124
OY 245 ProlYsProTrpGlnLeuProSerGlyValGlnProValGlnAlaLysLysSerArgile 264
DB 125 CCCAAGTCTTGGACGCTCCACATGCTGTGAGCCTCAAGTGCACATAGTCAAGAAATT 184
OY 265 GluValTrpGlnProProLleArgPheGlnLysLleIerGlyasnProTrpMetProArg 284
DB 185 GGAGTTTGAGAACTCCCTCCATTCAGAGGATGATGAATGCAATGCTGGATGTCAGG 244
OY 285 GlnLysPheAlaValGlnValGlySerSerTrpParThrSerAlaArgValValGlnLys 304
DB 245 AAGAGGTGCTACAGGGGCAAGGCACTCATGAGAACTCTGCTAGAGACAGTGCAGAAAG 304

OY 305 GlnYasnValGlyTrpGlnProPheHisArgValProSerGlyAlaProSerSerArgila 324
DB 305 GGAATATGTGGGTCCAGAGCCCCACACAGAGTCTCTATGTGGGCACTGCTAGTGAGGT 364
OY 325 ValArgArgSerProProSerSerArgLeuGlnLysGlyArgSerThrAspSerLeuGln 344
DB 365 ATCAGAAAGGGCCATCTACTCCGAGACCCAGCATGTGTATGCACACACACTTTCAC 424
OY 345 HisValProGlnLysSerThrAspThrGlnGlnGlnProValLysAlaAlaGlyMetGln 364
DB 425 TGTGCACCTGGAAAAAGCCACAGACACTCAATGCCAGCCATGAAGCCGTGAGAGGAG 484
OY 365 SerValProTrpLysThrValAlaGluLeuThrLysThrValGlyLleTrpLeuLeu 384
DB 485 GCTGTACCTTACAAAGCCACAGGACAGAGTCCGCCAAGCCATGTGAACCCACTTCTG 544
OY 385 HisCysHisAspLeuAspValArgHisGlyAlaLysArgAspHisPheGlyAlaLeuArg 404
DB 545 CATCAGCATGACTCGAGATGTGAGACATGAGTCAAGAGATATTTTGGAGCTTTAAG 604
OY 405 PheAspCysProThrGlyPheArgThrTrpMetGlyProVal-ProLeuCysPheGlygl 424
DB 605 TTTCAGTCCGCCACATGATTTGGACTTGCCTTCAGCCCTGTAGCCCTTTGTTGGCCA 664
OY 424 nphePhePro-PheGlyThrAlaValPheThrGlnCysLeuTrpLeuHisCysMet 442
DB 665 ATTTCACATGTGGGAATGCTGTATTACTCAATGCTGAAATCCCATTTGATC 720

RESULT 12
B0422247/c 1080 bp mRNA linear EST 23-MAY-2002
LOCUS B0422247
DEFINITION AGENCOURT_7766790 NIH_MGC_92 Homo sapiens cDNA clone IMAGE:6015590
5', mRNA sequence.
ACCESSION B0422247
VERSION B0422247.1 GI:21117562
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 1080)
AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabs-r@mail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: L1AM13211 row: o column: 15
High quality sequence stop: 702.
Location/Qualifiers

FEATURES
source
1. 1080
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:6015590"
/clone_id="NIH_MGC_92"
/tissue_type="embryonal carcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: testis; Vector: pCMV-SPORT6; Site:1: NotI;
Site:2: SalI; Cloned unidirectionally; oligo-dT primed.
Average insert size 2.5 kb. Library enriched for
full-length clones and constructed by Life Technologies.
Note: this is a NIH_MGC library."

BASE COUNT 239 a 304 c 221 g 316 t
ORIGIN
Alignment Scores: 2.53e-56 Length: 1080
Pred. No.: 2.53e-56

Score: 735.00 Matches: 161
 Percent Similarity: 71.60% Conservative: 18
 Best Local Similarity: 64.40% Mismatches: 56
 Query Match: 30.42% Indels: 16
 DB: 14 Gaps: 2

US-09-997-610-2_copy_18_459 (1-442) x BQ422247 (1-1080)

```

QY 101 AsnLierpleuMetArgLysGlnIleLeuAlaAsnLysGluGlnIleSerLysGlnGln 120
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 746 AACCTTGAACCTTGAGAGATGATTTAGAGTACTGCTGAGAGAGATTCTTAAGACGAA 687
QY 121 SerIleGlnIleValThrTrpValLeuLeuLysAlaPheSerPheIleArgGluIleVal 140
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 686 AGCATTCAGAGGTGACTTGGTGGCTGTTAAAGCATTCAGTTTGTAAAGGAAACAGAG 627
QY 141 HisLysSerSerGluAsnLeuHisProAspAsnValIleLysLysLysAsn-ProPheSe 160
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 626 CATAAAGATTCAAAAATTTGACGCTGACATGCAATGAAAGAAAATCTCATTTTC 567
QY 160 rGluGlyLysPheLysLeuAlaIleGluIleCysIleCysAsnGluGluLeuAsnValAs 180
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 566 TGAGGAGAAATGAGAGAGCTGCAAGAAATCGCTAAGTAAAGAAAGCCAAATGTTAA 507
QY 180 nProGlnAspAsnGlyLysAsnIleSerTrpThrCysGlnArgSerSerGlnGlnSerIl 200
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 506 TCCTCAAGACACAGGGGAAATGCTCCAGGATATGTACAGATT-TCATGGCAGCCCT 448
QY 200 eLysSerLeuAlaTrpArgPro--ArgArgLysTrpPheCysGlyThrGlyPro-Glys 219
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 447 TCCATCACAGGCCAGAGGCTCAGCAAGAAAGTGTCTGCGCCAGGCCAGGGTA 388
QY 219 erLeu-CysCysValGlnProArgAspLeuValProCysValProValAsn----- 235
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 387 CACTGTGCTGTGTGACGACCAAGAGACTGTGCTGTCCATCCAGCTCCAGACATG 328
QY 236 -----SerAlaValAlaSerGluGlyAlaSerProLysPro 247
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 327 GCTGAAGAGGCCCAACATAGAGTTCCGGCAGTGCTTACAGAGGTGCAAGGCCCAAGCCT 268
QY 248 TrpGlnLeuProSerGlyValGluProValGlyAlaLysLysSerArgIleGluValTrp 267
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 267 TGGCAGCTTCATGTGAGTGTGAGCCTGAGAGTGACAAAGTCAAAAGTGGGGTTGG 208
QY 268 GluProProIleArgPheGlnLysIleLysGlyAsnProTrpMetProArgGlnLysPhe 287
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 207 GAACCTCTGCTCAAGATTCTGAGAGTGTAGAAATGCTGATGCTCAGGCTAAGTTT 148
QY 288 AlaValGlyValGlySerSerTrpArgThrSerAlaArgValIleGlnLysGlyAsnVal 307
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 147 GCTGACAGGGGAGGGCTTCATGGCGAACCACCTGTTAGGGCAGTGGAGAAAGAAATGTG 88
QY 308 GlyTrpGluProProHisArgValProSerGlyAlaProSerSerArgAlaValArgArg 327
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 87 AGGTCAGAGCCCCACACAGAGTTCCCTGGGGCACCACCTAGTGGAGCTGTGAAGA 28
QY 328 SerProProSerSerArgLysGln 335
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 27 GGGCCACTGTCTCCAGACCTCAG 4

```

```

AUTHORS Poustka,A., Wellenreuther,R., Mewes,H.W., Well,B. and Wiemann,S.)
TITLE EST (Poustka,A., Wellenreuther,R., Mewes,H.W., Well,B. and Wiemann
JOURNAL Unpublished (1999)
COMMENT Contact: Poustka A.J.
          Department Lehrach
          Max-Planck-Institute for Molecular Genetics
          Ihnestrasse 73, 14195 Berlin, Germany
          Tel: +49-30-84131623
          Fax: +49-30-84131128
          Email: poustka@mpimg-berlin-dahlem.mpg.de
          This is the 5' sequence of the clone insert
          Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
          Research Center (DKFZ), Email s.wiemann@dkfz-heidelberg.de;
          sequenced by DKFZ (German Cancer Research Center,
          Heidelberg/Germany) within the cDNA sequencing consortium of the
          German Genome Project.
          No sl sequence available.
          This clone (DKFZp686B0612) is available at the RZPD in Berlin.
          Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
          Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.
FEATURES
source Location/Qualifiers
1..699
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="DKFZp686B0612"
/clone_lib="686 (synonym: hlc3)"
/tissue_type="human skeletal muscle"
/dev_stage="adult"
/lab_host="DH10B"
/note="Vector: pTriplex2; site_1: sfIla; site_2: sfIIB;
CDNA-collection"
BASE COUNT 147 a 206 c 171 g 175 t
ORIGIN
Alignment Scores:
Pred. No.: 1,99e-56 Length: 699
Score: 733.00 Matches: 152
Percent Similarity: 70.69% Conservative: 12
Best Local Similarity: 65.52% Mismatches: 56
Query Match: 30.34% Indels: 14
DB: 9 Gaps: 1
US-09-997-610-2_copy_18_459 (1-442) x AL602279 (1-699)
QY 208 ArgArgLysTrpPheCysGlyThrGlyProGlySerLeuCysValGlnProArgasp 227
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 691 AGAAAAAATGTTTCATGTGAGGCCAAACCAGAGGTCCCTGCTGTGGCCAGTCTAGGGAC 632
QY 228 leuValProCysValProValAsnSerAlaValAlaSer----- 240
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 631 TTGTGCTCTGATCCTCAGCCACCTAGCTGTGACCGAAAGGGCCCAATGTAGAAATTCAG 572
QY 241 -----GluGlyAlaSerProLysProTrpGlnLeuProSerGlyValGluPro 256
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 571 GCCATGTCTCCAGAGGGTCAAGCTCAAGCCTTGGCAGCTTCCAGTGGTGTAGAGCT 512
QY 257 ValGlyAlaLysLysSerArgIleGluValTrpGluProProIleArgPheGlnLysIle 276
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 511 GTGAGTGCATAGAGTCAAGATCAAGAGTGTGGAAACCTCTCCTAAGATTTCAGAGGGTG 452
QY 277 TyrGlyAsnProTrpMetProArgGlnLysPheAlaValGlyValGlySerSerTrpArg 296
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 451 TATGGAACACCTGGATGCCAGGAGAAATGTTGCTGCGAGGGCAGAGCCTTCATGAGAGA 392
QY 297 ThrSerAlaArgValValGlnLysGlyAsnValGlyTrpGluProProHisArgValPro 316
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 391 ACCTGTGCTAGGGCAGTGCAGAAAGGAAATGCGGGTCAAGACCCCAACAGAGATTTCC 332
QY 317 SerGlyAlaProSerSerArgAlaValArgArgSerProProSerSerArgLysGlnLys 336
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 331 ACTGAGGACACTGCTCAAGTGCAGCTATGAGAAAGGCGCATGTGCTCCAGACCCCAAGAT 272

```


TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabs-remail.nih.gov
Tissue Procurement: ATCC

CDNA Library Preparation: Life Technologies, Inc.
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LMN)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LMN at:
http://image.llnl.gov
Plate: LLM1214 Row: a Column: 13
High quality sequence stop: 599.
Location/Qualifiers

FEATURES

SOURCE

1. 998
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5492244"
/clone_1db="NIH_MGC_67"
/tissue_type="retinoblastoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: eye; Vector: PCMV-SPORE6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo
Average insert size 1.75 kb. Library constructed by Life
Technologies."

BASE COUNT 252 a 241 c 284 g 219 t 2 others
ORIGIN

Alignment Scores:

Pred. No.:	5.19e-56	Length:	998
Score:	731.00	Matches:	183
Percent Similarity:	64.65%	Conservative:	20
Best Local Similarity:	58.28%	Mismatches:	83
Query Match:	30.26%	Indels:	30
DB:	14	Gaps:	5

US-09-997-610-2_COPY_18_459 (1-442) x BM03785 (1-998)

QY 142 LYSSESSERGIUASNLEUHIIPROASPASNVALILELYSLYSASN--PROPHESERG 161
DB 2 AAAAGTTGGAAAATTTCACCTGCACAAATGTGATGAAAGAAAATTCCTATTTCTG 61
QY 161 LUGLYLSPHELYSLLEUALAAGLUILLIECYSLIECYASNGLUGLILEUASNVALASN 181
DB 62 AGGAGAAATTTAAGCTGGCTGCAGAAATTTGCATGATGACAG--GAGCCAAATGCTAATT 120
QY 181 ROGINSASPANGLYGLUASNLESERTPRTHCYSGLNARGSERSEGINLSERILEL 201
DB 121 CCCAAGACAAATGGGGAATATGCTCCAGGCGCATGTCCAGAGTCTTTATGGCAACCCCTCC 180
QY 201 YSSERLEUALTRPARG-PROARGARGLYSTRPHECYSGLYTHGLYPROGLYSERLEU 220
DB 181 CATCACAGCTCCAGATATACAGAAAATGTTTGTGGCCAGGCCGGGGTCTCA 240
QY 221 CYSCYSVALGINPROARGASPLEUVALPROCYVALPROVAL----- 234
DB 241 TGCTGTGTGACAGCTTAGGGACTGTGGCTGCATGCCAGCCACCTCCCAACCATGACTGA 300
QY 235 -----ASNSER-ALAVALLASERGLUGLYALASERPROLYSPROTRPGL 249
DB 301 CAGGAGGAGGTAGAGCTTGCGGCTGTAGCTTCGGGGAGTGCAGACCCCAACCTTGACA 360
QY 249 NLEUPROSERGLYVALGIUPTROVALGLYALALYSLSERARGILEGLUVALTRPGLUPR 269
DB 361 GCTTCCAGTGTGTGAGACTGCGAGTGCACAGAAAGTCAAGAACTGGGGTTTGGAAACC 420
QY 269 OPROLEARGPHEGLINLSYLIETRYGLYASNPROTRPMETPROARGGLINLSPHEALAVA 289
DB 421 TTCCGCTTAGATTAAAGAGGATGTGCGGAATGCCCTGGATGGCCAGGACGAAGTTTGTGTC 480
QY 289 GLYVALGLYSERSTRPARGTHTSERALARGVALVALGINLSGLYASNVALGLYTR 309
DB 481 AGGGGACAGGGCCCTCATGTGAGATCTCTGCCAGGCGCATGCGAAGGGAATGTGGGTC 540

QY 309 PGIUPROPHNISARGVALPROSERGLYALAPROSERSERARGALAVARGSERPR 329
DB 541 AGAGACCCACACACAGTCCCTACTGGGGACCCACCTAGTGAAGTGAAGAGAGTCC 600
QY 339 OPROSERARGLEUAGLINLSYGLYARGSERTHRASPERSERLEUGLINLSVALPROGLULY 349
DB 601 TCCAGACCC-----CAGAAATGTAGATCCACCGACAGCTTGACACCGTGTACGTGANA 653
QY 349 SSERTHRSPHRGINCYSGLINPROVALLYSLAALAGLYMETGLUSERVALPROTYR-L 369
DB 654 AGTTCACAGCACTNCATGCCACCCATGAAACAGCTGAGAGGAGGCTGTACCCCTGCA 713
QY 369 YSTHVALVALAAGLUUETHRLYSTHR-VALGLYLIETRY-LEULHNISCYSHNIS 388
DB 714 AGGTACAGGGGCGAGAGCTGCCCAAGACCATGGAGAACCCCTTNCATCTGCAATGA 773
QY 388 PLEUASPVALARGNISGLY-----VALYSARGASPHIS-PHEGLYALALEUARG-PHE 405
DB 774 CCTTGATGTGAANAATGGTGGAGTGCACAGAGGAGATCATTTTGTGAGCTTTTAGATTT 833
QY 406 ASPCYSPROTHR---GLYPHEARGTHRYMETGLYPRO-----VALPROLEUCYSRHE 422
DB 834 GACTGGCCCCCTGGGATTTGTGACTCCTCATGGGGCCTTGGAAACCTCTCTTGGTTTG 893
QY 423 GLYGLNPNH-PHEPROPHGLY 429
DB 894 ACCCAATTTATTCCTCATTTGGG 915

Search completed: February 19, 2003, 19:35:44
Job time : 3455.83 secs

THIS PAGE BLANK (USPTO)

Gencore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: February 19, 2003, 14:57:18 ; Search time 6158.57 Seconds

(without alignments)
2088.706 Million cell updates/sec

Title: US-09-997-610-2_COPY_18_459

Perfect score: 2416
Sequence: 1 AGPPAHPRPPEVGPARG.....GQFPFGNAFTQCILYHGM 442

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL-frame_p2n.model -DEV-xlh
-O/-cgn2.1/USPTO/spool/US09997610/runat_10022003_160823_23819/app_query.fasta_1.1635
-DB-GenEmbl -QFMT-fastap -SUFFIX-rge -MINMATCH=0.1 -LOOPEL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blomsun62 -TRANS=human40.cdi -LIST=45
-OCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -MODE-LOCAL
-OUTFMT=ptc -NORM-ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09997610.ecgn_1_1_3965-etunat_10022003_160823_23819 -NCPU=6 -ICPU=3
-NO_XLPHY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

GenEmbl:*
1: gb_da:*
2: gb_hg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*
15: em_da:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*

29: em_vl:*
30: em_hcg_hum:*
31: em_hcg_inv:*
32: em_hcg_other:*
33: em_hcg_mus:*
34: em_hcg_pin:*
35: em_hcg_rtd:*
36: em_hcg_mam:*
37: em_hcg_vrt:*
38: em_sy:*
39: em_hcgo_hum:*
40: em_hcgo_mus:*
41: em_hcgo_other:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	2327	96.3	145880	9	HS302D9	282198 Human DNA s
2	1218.5	50.4	161575	2	AC025233	AC025233 Homo sapi
3	1218.5	50.4	171625	2	AC087500	AC087500 Homo sapi
4	1217.5	50.4	123098	9	AC008799	AC008799 Homo sapi
5	1217.5	50.4	174874	9	AC123595	AC123595 Homo sapi
6	1217.5	50.4	189768	2	AC044889	AC044889 Homo sapi
7	1209	50.0	91323	9	AC106901	AC106901 Homo sapi
8	1207	50.0	112515	9	AL139216	AL139216 Human DNA
9	1205	49.9	137785	9	HS345P10	282201 Human DNA s
10	1200	49.7	168080	2	AC009047	AC009047 Homo sapi
11	1200	49.7	199579	9	AC093536	AC093536 Homo sapi
12	1199.5	49.6	155764	9	AC011238	AC011238 Homo sapi
13	1199.5	49.6	187898	2	AC016715	AC016715 Homo sapi
14	1198	49.6	99084	2	AC026420	AC026420 Homo sapi
15	1198	49.6	133780	9	AC010280	AC010280 Homo sapi
16	1198	49.6	170723	9	AC013470	AC013470 Homo sapi
17	1198	49.6	174293	9	AC008816	AC008816 Homo sapi
18	1198	49.6	174293	9	AC010242	AC010242 Homo sapi
19	1198	49.6	203433	9	AC017093	AC017093 Homo sapi
20	1195.5	49.5	186660	2	AC026107	AC026107 Homo sapi
21	1195	49.5	40714	9	HS0212C1	269722 Human DNA s
22	1194.5	49.4	124518	9	AL138962	AL138962 Human DNA
23	1194	49.4	166757	9	AC003046	AC003046 Homo sapi
24	1194	49.4	185171	2	AC023098	AC023098 Homo sapi
25	1193.5	49.4	174380	2	AC016075	AC016075 Homo sapi
26	1193.5	49.4	175559	9	AC112498	AC112498 Homo sapi
27	1193	49.4	68935	9	AC008150	AC008150 Homo sapi
28	1192	49.4	77405	2	AL353634	AL353634 Homo sapi
29	1192	49.4	176784	2	AC016632	AC016632 Homo sapi
30	1192	49.3	182126	2	AC016639	AC016639 Homo sapi
31	1190.5	49.3	154090	2	AC025577	AC025577 Homo sapi
32	1190	49.3	149080	2	AC087699	AC087699 Homo sapi
33	1187	49.1	168502	9	AC091005	AC091005 Homo sapi
34	1186.5	49.1	172919	9	AL807742	AL807742 Human DNA
35	1185.5	49.1	123631	9	HS22F01	AL109967 Homo sapi
36	1184.5	49.0	94730	9	AP000230	AP000230 Homo sapi
37	1184.5	49.0	100000	9	AP000144	AP000144 Homo sapi
38	1184.5	49.0	100634	9	AP001594	AP001594 Homo sapi
39	1184.5	49.0	182778	2	AC022312	AC022312 Homo sapi
40	1184.5	49.0	340000	9	AP001695	AP001695 Homo sapi
41	1184	49.0	173480	9	CNS00W6T	AL079343 Human chr
42	1183	49.0	159875	9	AC024033	AC024033 Homo sapi
43	1182.5	48.9	153223	2	AC074339	AC074339 Homo sapi
44	1182	48.9	110820	9	AL596285	AL596285 Human DNA
45	1181	48.9	182972	2	AC023550	AC023550 Homo sapi

RESULT 1

ALIGNMENTS

```
HS302D9          145880 bp      DNA      linear      PRI 12-DEC-1999
LOCUS            Human DNA sequence from clone RP1-302D9 on chromosome 22 Contains
DEFINITION       GSSs, complete sequence.
ACCESSION        282198
VERSION          282198.2  GI:6572207
KEYWORDS         HTG.
SOURCE           Homo sapiens.
ORGANISM         Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE        1 (bases 1 to 145880)
AUTHORS          Bridgeman,A.
TITLE            Direct Submission
JOURNAL          Submitted (08-DEC-1999) Sanger Centre, Hinxton, Cambridgeshire,
                CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
                Requests: clonerequest@sanger.ac.uk
                On Dec 13, 1999 this sequence version replaced gi:3164067.
                During sequence assembly data is compared from overlapping clones.
                Where differences are found these are annotated as variations
                together with a note of the overlapping clone name. Note that the
                variation annotation may not be found in the sequence submission
                corresponding to the overlapping clone, as we submit sequences with
                only a small overlap as described above.
                This sequence has been finished according to sequence map criteria
                as follows. An attempt is made to resolve all sequencing problems,
                such as compressions and repeats, but not necessarily within known
                annotated human repeat sequence elements (e.g. Alu). Where the
                sequence is ambiguous, there is an annotation using the 'unsure'
                feature key.
                The following abbreviations are used to associate primary accession
                numbers given in the feature table with their source databases:
                Em, EMBL; Sw, SWISSPROT; Tr, TrEMBL; Wp, WORMPEP; Information
                at http://www.sanger.ac.uk/Projects/C.elegans/wormpep This sequence
                was generated from part of bacterial clone configs of human
                chromosome 22, constructed by the Sanger Centre Chromosome 22
                Mapping Group. Further information can be found at
                http://www.sanger.ac.uk/HGP/Chr22
                RP1-302D9 is from the library RPC1-1 constructed at the Roswell
                Park Cancer Institute by the group of Pieter de Jong. For further
                details see http://bacpac.med.buffalo.edu/
                VECTOR: pCYPAC2
                This sequence is the entire insert of clone RP1-302D9 The true left
                end of clone CTA-282F2 is at 69682 in this sequence. The true right
                end of clone CTA-415G2 is at 55167 in this sequence.
                Location/Qualifiers
                1..145880
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /chromosome="22"
                /clone="RP1-302D9"
                /clone_id="RPC1-1"
                188..245
                repeat_region
                246..571
                repeat_region
                572..759
                repeat_region
                783..933
                repeat_region
                1033..1336
                repeat_region
                1450..1583
                repeat_region
                1687..1752
                repeat_region
                2350..2660
                repeat_region
                2684..2981
                repeat_region
                3323..3343
                repeat_region
                /note="MIR repeat: matches 144. 209 of consensus"
                /note="AluSc repeat: matches 1. 312 of consensus"
                /note="MIR repeat: matches 1. 144 of consensus"
                /note="MIR repeat: matches 26. 187 of consensus"
                /note="AluSc repeat: matches 1. 299 of consensus"
                /note="MIR repeat: matches 24. 160 of consensus"
                /note="AluSc repeat: matches 2593. 2661 of consensus"
                /note="AluSc repeat: matches 3. 309 of consensus"
                /note="AluSc repeat: matches 2. 300 of consensus"
                /note="MIR repeat: matches 116. 136 of consensus"

FEATURES
Source
1..145880
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="22"
/clone="RP1-302D9"
/clone_id="RPC1-1"
188..245
repeat_region
246..571
repeat_region
572..759
repeat_region
783..933
repeat_region
1033..1336
repeat_region
1450..1583
repeat_region
1687..1752
repeat_region
2350..2660
repeat_region
2684..2981
repeat_region
3323..3343
repeat_region
/note="MIR repeat: matches 144. 209 of consensus"
/note="AluSc repeat: matches 1. 312 of consensus"
/note="MIR repeat: matches 1. 144 of consensus"
/note="MIR repeat: matches 26. 187 of consensus"
/note="AluSc repeat: matches 1. 299 of consensus"
/note="MIR repeat: matches 24. 160 of consensus"
/note="AluSc repeat: matches 2593. 2661 of consensus"
/note="AluSc repeat: matches 3. 309 of consensus"
/note="AluSc repeat: matches 2. 300 of consensus"
/note="MIR repeat: matches 116. 136 of consensus"

repeat_region 3344..3652
/note="Alu repeat: matches 1. 309 of consensus"
repeat_region 3653..3928
/note="MIR repeat: matches 136. 359 of consensus"
repeat_region 3929..4278
/note="MIR repeat: matches 3. 364 of consensus"
repeat_region 4279..4485
/note="MIR repeat: matches 359. 568 of consensus"
repeat_region 5073..5176
/note="MIR repeat: matches 1. 311 of consensus"
repeat_region 5181..5491
/note="MIR repeat: matches 1. 311 of consensus"
repeat_region 6369..6485
/note="MIR repeat: matches 2579. 2705 of consensus"
repeat_region 6647..6685
/note="MIR repeat: matches 1. 23 of consensus"
repeat_region 6686..6987
/note="MIR repeat: matches 1. 302 of consensus"
repeat_region 6988..7036
/note="MIR repeat: matches 23. 77 of consensus"
repeat_region 7482..7754
/note="MIR repeat: matches 9. 290 of consensus"
repeat_region 7775..8060
/note="MIR repeat: matches 1. 295 of consensus"
repeat_region 8414..8551
/note="MIR repeat: matches 2553. 2706 of consensus"
repeat_region 8914..9030
/note="MIR repeat: matches 147. 262 of consensus"
repeat_region 9110..9280
/note="MIR repeat: matches 91. 262 of consensus"
repeat_region 9283..9412
/note="MIR repeat: matches 15. 144 of consensus"
repeat_region 9521..9679
/note="MIR repeat: matches 3. 161 of consensus"
repeat_region 9920..10225
/note="MIR repeat: matches 2. 425 of consensus"
repeat_region 10179..10678
/note="MIR repeat: matches 2. 425 of consensus"
repeat_region 10718..11310
/note="MIR repeat: matches 1. 144 of consensus"
repeat_region 10784..11201
/note="MIR repeat: matches 1. 144 of consensus"
repeat_region 11838..11946
/note="MIR repeat: matches 20. 137 of consensus"
repeat_region 12174..12445
/note="MIR repeat: matches 1988. 2275 of consensus"
repeat_region 12444..12642
/note="MIR repeat: matches 63. 241 of consensus"
repeat_region 13017..13369
/note="MIR repeat: matches 174. 244 of consensus"
repeat_region 13331..13397
/note="MIR repeat: matches 1. 302 of consensus"
repeat_region 13398..13698
/note="MIR repeat: matches 1. 302 of consensus"
repeat_region 13699..13810
/note="MIR repeat: matches 76. 174 of consensus"
repeat_region 13806..13919
/note="MIR repeat: matches 77. 189 of consensus"
repeat_region 13945..14060
/note="MIR repeat: matches 24. 142 of consensus"
repeat_region 14061..14367
/note="MIR repeat: matches 1. 301 of consensus"
repeat_region 14368..14452
/note="MIR repeat: matches 141. 225 of consensus"
repeat_region 14589..14679
/note="MIR repeat: matches 173. 262 of consensus"
repeat_region 14597..15201
```

```

mlsc-feature      /note="match: GSS: Em: A0553482"
14616..15060
repeat_region     /note="match: GSS: Em: A0370601"
14868..15040
repeat_region     /note="MIR repeat: matches 49..233 of consensus"
15071..15188
repeat_region     /note="I2 repeat: matches 2112..2239 of consensus"
15304..15399
repeat_region     /note="MTIB repeat: matches 1..99 of consensus"
15490..15662
repeat_region     /note="AluSg1 repeat: matches 2..114 of consensus"
15669..15727
repeat_region     /note="MTIB repeat: matches 119..178 of consensus"
15728..16027
repeat_region     /note="AluSc repeat: matches 1..299 of consensus"
16028..16245
repeat_region     /note="MTIB repeat: matches 178..390 of consensus"
16546..16854
repeat_region     /note="AluY repeat: matches 1..300 of consensus"
18296..18323
repeat_region     /note="MSTA repeat: matches 2..29 of consensus"
18324..18392
repeat_region     /note="MER66-Internal repeat: matches 4919..4993 of consensus"
18393..18712
repeat_region     /note="Aluub repeat: matches 1..311 of consensus"
18713..19133
repeat_region     /note="MER66-Internal repeat: matches 4548..4919 of consensus"
complement(18872..19230)
mlsc-feature      /note="match: GSS: Em: A0005063"
19251..19719
repeat_region     /note="match: GSS: Em: B14179"
19537..20290
repeat_region     /note="HERVH21 repeat: matches 4657..5784 of consensus"
20317..20382
repeat_region     /note="33 copies 2 mer ta 68 conserved"
20513..20666
repeat_region     /note="77 copies 2 mer tt 70 conserved"
20682..21008
repeat_region     /note="AluSg1 repeat: matches 1..306 of consensus"
21239..21553
repeat_region     /note="HMER8-P3 repeat: matches 4410..4713 of consensus"
21882..22254
repeat_region     /note="THE1B repeat: matches 1..364 of consensus"
22302..22537
repeat_region     /note="MER66-Internal repeat: matches 2186..2417 of consensus"
22538..22850
repeat_region     /note="AluSP repeat: matches 1..313 of consensus"
22851..23801
repeat_region     /note="MER66-Internal repeat: matches 1210..2186 of consensus"
23905..23989
repeat_region     /note="MER66-Internal repeat: matches 3017..3102 of consensus"

Alignment Scores:
Pred. No.:      1.6e-116      Length:      145880
Score:          2327.00      Matches:      427
Percent Similarity: 100.00%      Conservative: 0
Best Local Similarity: 100.00%      Mismatches: 0
Query Match:    96.32%      Indels:      0
DB:             9           Gaps:         0

US-09-997-610-2_COPY_18_459 (1-442) x HS302D9 (1-145880)
OY 16 ProglyAlaProglyLeuProglyIleuProglyIleuSerGluMetThrIleuSerGlyPro 35
|||||
DB 37262 CCGGGGCCCCAGGTTTACACACATATACGCGGAATTAAGTAAATGACAAATGCCCC 37321

OY 36 CysProAspIleGluArgSerAlaPheThrValLysLeuSerGlyLysLeuProLeuPro 55
|||||
DB 37322 TGTCGTGATATAGAAAGTCAGCCTTTACTGTGACACTGACGAAACACTCTCTTCC 37381

```

```

OY 56 PheLysProIleIlePheThrGlyValIleuTyAsnAlaGlnArgAspLeuLysGluVal 75
|||||
DB 37382 TTCAAGCCCATCATCTTCCACAGGGGCTCTGTACATATCCCAAGAGGATTTAAAGAGGCC 37441

OY 76 MetGlyValPheAlaCysArgValProGlyAsnTyTrpSerSerPheAspValGluLeu 95
|||||
DB 37442 ATGGAGCTTCTTCTTCAGAGGTCCTCGGAATTAATCAACACTTGAATGTTGACCTG 37501

OY 96 HisIscLysLysValAsnIleTrpLeuMetArgLysIleIleuAlaAsnLysGluGlu 115
|||||
DB 37502 CATCATTCAGAGGATTAATTTGGCTAATGAGCAAGCAAAATTTGGCTAATGAGGAA 37561

OY 116 IleSerLysGlnSerIleGlnGluValThrTrpValIleuLeuLysAlaPheSerPhe 135
|||||
DB 37562 ATTTCTAAGCAGCAAGACATTCAGAGGTGACTTGGTCTGTAAAGCATTCAGTTTC 37621

OY 136 IleArgGluAlaGluHisLysSerSerGluAsnLeuHisProAspAsnValIleLysLys 155
|||||
DB 37622 ATAAAGGAGGACAGCATTAAGAGTTGAGAAATTTGCACCTGCACATATGATGATGAG 37681

OY 156 LysAsnProPheSerGluGlyLysPheLysLeuAlaAlaGluIleCysIleCysAsnGlu 175
|||||
DB 37682 AAAAACCATTCTTGAGGGGAAATTCACCTGCTGCACAAATTTGCATATGATGAG 37741

OY 176 GluLeuAsnValAsnProGlnAspAsnGlyGluAsnIleSerTrpThrCysGlnArgSer 195
|||||
DB 37742 GAGCTGATTTATCTTCAAGACATAGGGGAAATTTCTCTGACATGTACAGAGTCT 37801

OY 196 SerGlnSerIleLysSerLeuAlaTrpArgProArgArgLysTrpPheCysGlyThr 215
|||||
DB 37802 TCACAGCAGTCATCAATATCAGGCTGAGGCTGAGGAAATGTTTGGGGACA 37861

OY 216 GlyProGlySerLeuCysCysValGlnProArgAspLeuValProCysValProValAsn 235
|||||
DB 37862 GGGCCAGGTCCTGTGCTGTGTGCACCTAGAGACTTGTGCTGTGCCAGTTAAT 37921

OY 236 SerAlaValAlaSerGluGlyAlaSerProLysProTrpGlnLeuProSerGlyValGlu 255
|||||
DB 37922 TCACGTGTGCTTCAGAGGGGTGCAGGCCCAAGCCTTGGAGCTTCAATGAGTTGAG 37981

OY 256 ProValGlyAlaLysLysSerArgIleGluValTrpLeuProProIleArgPheGlnLys 275
|||||
DB 37982 CCGTGGGTGCAAAAGATGACAGATTTGGGAACTTCACATCAGATTTTCAAGAG 38041

OY 276 IleTyGlyAsnProTrpMetProArgGlnLysPheAlaValGlyLysSerTrp 295
|||||
DB 38042 ATATATGAAACCCCTGATGCCAGGCAAGATTTGCTGTAGGGGTGGCTCTCATGG 38101

OY 296 ArgThrSerAlaArgValValGlnLysGlyAsnValGlyTrpGlnProPheArgVal 315
|||||
DB 38102 AGAACCTTCAGAGGATGAGTACAAAGGAAATGTTGGTGGAGCCCAACAGAGTGC 38161

OY 316 ProSerGlyAlaProSerSerArgAlaValArgArgSerProProSerSerArgLeuGln 335
|||||
DB 38162 CCCAGTGGGCTCATTTAGTATAGAGCTGTGAGAAAGATCCACCATCTCCAGACTCCAG 38221

OY 336 LysGlyArgSerThrAspSerLeuGlnHisValProGlyLysSerThrAspThrGlnCys 355
|||||
DB 38222 AAGGATGATCCACTGCACACTTGCAGCATGTGCTGAAAAATCCACAGACTCACTGC 38281

OY 356 GlnProValLysAlaAlaGlyMetGluSerValProTyTrpLysThrValValAlaGluLeu 375
|||||
DB 38282 CACCTGTGAAACAGCAGAGGATGAGATGTCGTACCTTCAACAAACCCGATGAGAGACTG 38341

OY 376 ThrLysThrValGlyIleTyLeuLeuHisCysHisAspLeuAspValArgHisGlyVal 395
|||||
DB 38342 ACCAAGACCTGGGAATCTTACCTTTCATTTGATGACTGAGACGTGAGACTGAGATC 38401

OY 396 LysArgAspHisPheGlyAlaLeuArgPheAspCysProThrGlyPheArgThrTyMet 415
|||||
DB 38402 AAAAGAGATCATTTTGGAGCTTTAAGATTTGAGCTGCCCCCAGCTGATTTGAGACTTATAG 38461

```

QY	416	GLYProValProIenueysphedylglnhphheprophegylmAlaValpMetArgln	435
Db	38462	GGGGCCGGTACCCCTTGTGTGTTTG6CCAAATTGTTCCATTGTGAACCTGCCGATTATTAACCAA	38521
QY	436	CysLeuTyrldeuHisCysMet	442
Db	38522	TGCGCTGTACCTCCATTGTATG	38542
RESULT 2			
AC025233/c			
LOCUS	AC025233	161575 bp	DNA linear HTG 07-JUL-2000
DEFINITION	Homo sapiens chromosome 17 clone RP11-333e1, WORKING DRAFT		
SEQUENCE	AC025233	27 unordered pieces.	
ACCESSION	AC025233	4	GI:8954317
VERSION	HTG; HTGS-PHASE1; HTGS-DRAFT.		
KEYWORDS	Homo sapiens.		
SOURCE	Homo sapiens		
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
REFERENCE	1 (bases 1 to 161575)		
AUTHORS	Waterston, R.H.		
TITLE	The sequence of Homo sapiens clone		
JOURNAL	Unpublished		
REFERENCE	2 (bases 1 to 161575)		
AUTHORS	Waterston, R.H.		
TITLE	Direct Submission		
JOURNAL	Submitted (07-MAR-2000) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA		
COMMENT	On Jul 7, 2000 this sequence version replaced gi:7801489.		

```

Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
Project Information -----
Center project name: H.NH0333E01
Summary Statistics -----
Sequencing vector: M13: 90%
Sequencing vector: plasmid: 10%
Chemistry: Dye-primer ET; 81% of reads
Chemistry: Dye-terminator Big Dye; 19% of reads
Assembly program: Phrap; version 0.990119
Consensus quality: 144558 bases at least Q40
Consensus quality: 149743 bases at least Q30
Consensus quality: 152625 bases at least Q20
Insert size: 176000; agarose-fp
Insert size: 158975; sum-of-contigs
Quality coverage: 3.48 in Q20 bases; agarose-fp
Quality coverage: 3.75 in Q20 bases; sum-of-contigs

NOTE: This is a 'working draft' sequence. It currently
* consists of 27 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1
1141: contig of 1141 bp in length
1142 1341: gap of unknown length
1142 1341: gap of unknown length
1242 2424: contig of 1183 bp in length
2425 3524: gap of unknown length
2525 3561: contig of 1437 bp in length
3562 4061: gap of unknown length
4062 5712: contig of 1651 bp in length
5713 5812: gap of unknown length
5813 8019: contig of 2207 bp in length
8020 8119: gap of unknown length
8120 10208: contig of 2089 bp in length
10209 10308: gap of unknown length
10309 11734: contig of 1426 bp in length

```

*	11735	11834	gap of unknown length
*	11835	13265	contig of 1431 bp in length
*	13266	13363	gap of unknown length
*	13366	14425	contig of 1060 bp in length
*	14426	14525	gap of unknown length
*	14526	15609	contig of 1084 bp in length
*	15610	15710	gap of unknown length
*	15710	17529	contig of 1820 bp in length
*	17530	17629	gap of unknown length
*	17630	20351	contig of 2722 bp in length
*	20352	20451	gap of unknown length
*	20452	23753	contig of 3302 bp in length
*	23754	23853	gap of unknown length
*	23854	26537	contig of 2684 bp in length
*	26538	31638	gap of unknown length
*	31638	31339	contig of 4702 bp in length
*	31440	31439	gap of unknown length
*	31440	37041	contig of 5602 bp in length
*	37042	37141	gap of unknown length
*	37142	45368	contig of 8127 bp in length
*	45369	45368	gap of unknown length
*	45569	52323	contig of 6955 bp in length
*	52324	52423	gap of unknown length
*	52424	60239	contig of 7818 bp in length
*	60240	60339	gap of unknown length
*	60340	68385	contig of 8046 bp in length
*	68386	68485	gap of unknown length
*	68486	78834	contig of 10349 bp in length
*	78835	78934	gap of unknown length
*	78935	90422	contig of 11488 bp in length
*	90423	90522	gap of unknown length
*	90523	100763	contig of 10241 bp in length
*	100764	100863	gap of unknown length
*	100864	111672	contig of 10809 bp in length
*	111673	111772	gap of unknown length
*	111773	126855	contig of 15083 bp in length
*	126856	126955	gap of unknown length
*	126956	142557	contig of 15602 bp in length
*	142558	142657	gap of unknown length
*	142658	161575	contig of 18918 bp in length

BASE COUNT	44288 a	35882 c	35114 g	43686 t	2605 others
ORIGIN					

Alignment Scores:

Pred. NO.:	6.53e-56	Length:	16155
Score:	1218.10	Matches:	28
Percent Similarity:	74.17%	Conservative:	239
Best Local Similarity:	66.39%	Mismatches:	77
Query Match:	50.43%	Indels:	16
DB:	2	Gaps:	3

US-09-997-610-2_COPY_18_459 (1-442) x AC025233 (1-161575)

QY 97 HiscysLysValAsnIleTrp-----LeuMetArgLysGlnIleLeuAlaAsnLys 113

Db 121671 CACTGCCCTAGAGATTGTGGAACTTTGAACTTGAGAGAGATGATTAGGTAATCTGGTG 121612

QY 114 GluIleSerLysGlnInsertIleGlnIleValThrTrpValLeuLeuLysAlaPhe 133

Db 121611 GAAGAAATTCTAAGCAGCAAAACATTCAAGGTGTGACTTGGGTCCTGTTAATGGATTC 121552

QY 134 SerpHe1eArg1uAlaGluHisLysSerSerGluAsnLeuHisProAspAsnValIle 153

Db 121551 AGTTTATGAGGGAACGAGCATAAAGTTCAGAAATTTCAGCCTGAGAATGTGATA 121492

QY 154 LysLysLysAsnProPheSerGluGlyLysPheLysLeuAlaAlaGluIleCysIleCys 173

```

DB 121491 GAAAGAAAAACCCATTTCTGTGAGGAGAAATTCATCAGCGCTGCACAAATTTTCATTAAGT 121432
OY 174 AaNgLuGlULeuaNsValaSnProGlnaSPaSnGlyGLuaSn1leSeRtrPrThrCysGln 193
DB 121431 AACCAAGAACCTGAATGTTAATCCGCAAGACAATGGGAAATGTCATCCAGGGCATGTCAg 121372
OY 194 ArgSeRSeGlnGlnSerIleLysSerLeuAlaTPaRgProARgArGlySTrPheCys 213
DB 121371 AAGTCTTCACACGAGCCCTCCATCCATCAGAGCCCTGGAGGCCCTAAGGAAATGCTTTTGT 121312
OY 214 G1YThrG1YProG1YSerLeuCYsCYsValGlnProARgASpleuValProCYsValPro 233
DB 121311 GGGCCAGGAGCCGAGGCTCCCAATGCTGTGTCCAGACCTAGGAGACTTGGTGGCTGCATCCCA 121252
OY 234 ValASerAlaValAlaSerGlnGly----- 242
DB 121251 GCGCGCTGATCCATGGCTGAAAGGGCCAAATGATAGACGTCACAGCTGTGGCTTCAGA 121192
OY 243 AlASerProLYsProTrpLInLeuProSerGlyValG1YProValG1YAlaLysLysSer 262
DB 121191 GGGTGGCCCTTAAGCCTTGGCAGATTTCACATGGTGTGATCTTTAGCTACACAGAAAGTCA 121132
OY 263 ArgG1leGlnValaTrpGlnProTrpIleArgpHeGlnLysIleGlyG1YAsnProTrpMet 282
DB 121131 AGAATTGAGGTTTGGGAGACTTCCACCTGATTTTCAGAAAGATGTATGAAACGCTTGATG 121072
OY 283 ProATGlnLYsPheAlaValG1YAlaG1YSerSerTrpArgTrpSerAlaArgValaVal 302
DB 121071 CCCAGCCAGAAATTTTCTCTCAGAGGGGAGGAGATCTCATGAGAAACCTCTCTAGAGGACATG 121012
OY 303 G1NLYsG1YAsnValG1YTrpG1YProPheAlaArgValaProSerG1YAlaProSerSer 322
DB 121011 CAGAAGGAGAAATGTGGGGTCGAAAGCCACCATATGAGTCCCTACTGCGGACACTTCTAGT 120952
OY 323 ArgAlaValaAlaArgArSerProProSerSerArgLeuGlnLysG1YARSerThrAspSer 342
DB 120951 GGAGCTGTGAAGAAGAGGGGCTGACCTTCCACACCCCAAAATGTATGATCCACCTACAGC 120892
OY 343 LeuGlnH1sValaProG1YAsnValaArgpHeGlnLysG1YProValaLysAlaValaGly 362
DB 120891 TTGCACACCGATATGCTGTGAAAGAGCTGCAGACACTCAACACCCAGCCCTGAAAGAGCCAGG 120832
OY 363 MetG1uSerValaProTrpLysThrValaValaG1YLeuThrLysThrValaGlyLeTyr 382
DB 120831 AGGAGAGGCTGTGCTCCCTACAAAGCCACAGGAGGTGAGAGTGTCCCAAGACCATGTGAAACCCAC 120772
OY 383 LeuL1uH1sCYsH1sAspLeuAspValaArgH1sG1YValaLysArgAspH1sPheG1YAla 402
DB 120771 CTCTTGACATCAGCATGACCTGTGATGTGAGAGCTGTGAAAGAGATATTTTGGAGGT 120712
OY 403 LeuATgPheASpCYsProThrG1YpHeArgTrpThrTyrMetG1YProVal-ProLeuCYsPh 422
DB 120711 TTAAATTTTGTATTTGCTTCTGTGATTTTGGAGCTTGCATGTGGCCCTGTAAACCCCTTTGT 120652
OY 422 eG1YlInPhePheProPheG1YThrAlaValaPheThr---G1NcyS1eutyR1euth1s 440
DB 120651 TGGCCAATTTCTCCATTGTGAAACAGCTGTATTATACCCAGCAATACCTGTACCCCAT 120594
RESULT 3
AC087500 171625 bp DNA linear HTG 13-AUG-2002
LOCUS Homo sapiens chromosome 17 clone RP11-333E1 map 17, *** SEQUENCING
IN PROGRESS ***, 2 ordered pieces.
AC087500
AC087500.15 GI:22213294
KEYWORDS HTG; HTGS_PHASE2; HTGS_FULITOP; HTGS_ACTIVEFIN.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 171625)
AUTHORS Bliren,B., Nusbaum,C. and Lander,E.
TITLE Homo sapiens chromosome 17, clone RP11-333E1

```

```

JOURNAL Unpublished
REFERENCE 2 (bases 1 to 171625)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,S.,
Barna,N., Bastien,V., Boguslavsky,L., Bokhgalter,B., Brown,A.,
Cammarata,J., Campopiano,A., Choepel,Y., Colangelo,M., Collins,S.,
Collamore,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S.,
Dodge,S., Faro,S., Ferreira,P., Fitzhugh,W., Gage,D., Galagan,J.,
Gardyna,S., Ginde,S., Goyette,M., Graham,L., Grand-Pierre,N.,
Hagos,B., Heatford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
Jones,C., Karatas,A., Larocque,K., Lamazares,R., Landers,T.,
Lehoczky,J., Levine,R., Liu,G., Maclean,C., Macdonald,P.,
Marquis,N., Matthews,C., McCarthy,M., McEwan,P., McKernan,K.,
McPheeters,R., Meldrim,J., Menus,L., Mihova,T., Mlenga,V.,
Murphy,T., Naylor,J., Nguyen,C., Norbu,C., Norman,C.H.,
O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K.,
Punkhang,P., Pierre,N., Pollara,V., Raymond,C., Retta,R.,
Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J., Rosetli,M.,
Roy,A., Santos,R., Schauer,S., Schnupack,R., Seaman,S., Severy,P.,
Sougnuez,C., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Strauss,N., Subramanian,A., Talamas,J., Testaye,S., Theodore,J.,
Travers,M., Travis,N., Triglio,J., Vassiliev,H., Viel,R., Vo,A.,
Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J.,
Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (05-JAN-2001) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 171625)
REFERENCE
AUTHORS Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
Barna,N., Bastien,V., Bloom,T., Boguslavsky,L., Bokhgalter,B.,
Cammarata,J., Chang,J., Chazaro,B., Choepel,Y., Collamore,A.,
Cook,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
Faro,S., Ferreira,P., Fitzgeraid,M., Gage,D., Galagan,J.,
Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hagos,B.,
Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A.,
Karatas,A., Kellis,C., Landers,T., Levine,R., Lindblad-Toh,K.,
Liu,G., Maclean,C., Macdonald,P., Major,J., Matthews,C.,
McCarthy,M., Meldrim,J., Menus,L., Mihova,T., Mlenga,V.,
Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H.,
O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K.,
Phunkhang,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P.,
Roman,J., Roy,A., Schauer,S., Schnupack,R., Seaman,S., Severy,P.,
Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talamas,J.,
Testaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J.,
Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (13-AUG-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Aug 13, 2002 this sequence version replaced gi:22122948.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/xw/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: W1BR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L12015
Center clone name: 333_E_1
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 2 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submittor.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
* 1 27961: contig of 27961 bp in length
* 27962 28061: gap of 100 bp
* 28062 171625: contig of 143564 bp in length.

```


	D	54561	GGGCCAGGCGCCAGAGGCTTTGTGTGTGCAGCTTAGGACATTGGGCCCTCATGCCA	54620
OY	234	Val	-	-asnsrAlaValAlaSerGIugly 242
D	54621	GTGTCTCCAGCACCATGGCTGAAGGGGCCAACATAGAGCTCGCGCTGTGGCTTCAGAAGT	54680	
OY	243	AlaseProIysProTrpGlnLeuProSerGIyValGluProValGIyAlaLysSyr 262		
D	54681	GCAAGCCCCAGGCTTGCGACGCTTCATCTGTGTGTAGCCTGCANAOTGCACGAAAGTCA	54740	
OY	263	ArgIIeGIuValTrpGluProProIIearPhneGInLysIleTryGIyAsnProTrpMet 282		
D	54741	CAAAATGGGGTTTGGGAACCTCCACCTAGATTGCAGATGATGTATGGAATGGCCGGATG	54800	
OY	283	ProArgGInLysPheAlaValGIyValGlySerSerTrpArgThrSerAlaArGIalVal 302		
D	54801	ACCAGAGAGACTTCTCTCACGGGACAGGCGCTTCATAGGAAACCTCTGCTAGGCGCATG	54860	
OY	303	GInLysGIyAsnValGIyTrpGluProProHIsArgValProSerGIyAlaProSerSer 322		
D	54861	CAGACAAAAATTTAGGCTCAGAGCCCACACACAGAGACCCACTAGAGCATCTGCTAG	54920	
OY	323	ArgAlaValArgTrpSerProProSerSerArgLeuGInLysGIyArgSerThraspSer 342		
D	54921	GGAGCTGTGGAAGAAGAGCCACCATCTCCAGATCCGAAATGSTAGTCCACCAACAG	54980	
OY	343	LeuGInHisValProGIyLysSerThrasPhrGIncysGInProValLysAlaGIy 362		
D	54981	TTCGACTGTGTACTGTAAGAAAACCCACAGAGCATCATTCCTCCAGCTGTGAAGACGACG	55040	
OY	363	MeGIuSerAlProTryLysThrValValAlaGluLeuThrLysThrValGIyIleTry 382		
D	55041	AGGAGAGCTATCCCTCGCAAAATCATACAGGGGCGAAGCTGTCCAAGACCTTGGGAACCTC	55100	
OY	383	LeuLeuHIScyHisHisPheLysPValArgHisGIyValLysAsnAspHisPheGIyAla 402		
D	55101	CTCTGCGCTCAGCGTACACCGGATGTGACATGTGACATGACATCAAGACATCTTTGGAGCT	55160	
OY	403	LeuArgPheaSPCystrOthrGIyPhearGThrTyrmelGIyProValProLeuCysPhe 422		
D	55161	TTTAAGATTGTACTGCCCTCTGGATTTCAGACTGAGTGGGCGCTGTAGCCCCCTTTGCTTT	55220	
OY	423	GIyGInPhePheProPheGIyThrAlaValPheThrGIncysLeuTryLeuHIScys 441		
D	55221	GGCCATTTCTCCATTGTGGAACAGCTGTATTACCCCATCCGTACCCCGCATGTGT	55277	
	RESULT 6			
	AC044889/c			
	LOCUS	AC044889	189768 bp	DNA linear HMG 24-AUG-2002
	DEFINITION	Homo sapiens chromosome 5 clone RP11-792Cl map 5, WORKING DRAFT		
	VERSION	AC044889		
	ACCESSION	AC044889.2	GI:8016676	
	KEYWORDS	HMG; HMG5_PHASE1; HMG5_DRAFT.		
	SOURCE	human.		
	ORGANISM	Homo sapiens		
	REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
	AUTHORS	Mammalia; Euthera; Primates; Catarrhini; Hominiidae; Homo.		
	JOURNAL	Birren,B., Linton,L., Nusbaum,C. and Lander,E.		
	TITLE	Homo sapiens chromosome 5, clone RP11-792Cl		
	REFERENCE	unpublished		
	AUTHORS	2 (bases 1 to 189768)		
		Birren,B., Baldwin,J., Barra,N., Bastien,V., Bede,F.,		
		Anderson,S., Baldwin,J., Brown,A., Burkett,G.,		
		Boguslavsky,L., Boukhalter,B., Brown,A., Buckell,G.,		
		Campioneiro,A., Castle,A., Choeheli,Y., Colangelo,M., Collins,S.,		
		Collymore,A., Cooke,P., Deairillano,K., Dewar,K., Diaz,J.S.,		
		Dodge,S., Domingo,M., Doyle,M., Ferreira,P., Fitzhugh,W., Gage,D.,		
		Galagan,J., Gardyna,S., Glinde,S., Goylete,P., Graham.L.,		
		Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,		
		Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kinn,L., Karatas,A.,		
		Klein,D., LaRoque,K., Lamazares,R., Landers,T., Lehoczy,J.,		

TITLE	JOURNAL	REFERENCE	AUTHORS
Submitted (12-APR-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA (bases 1 to 189768)			Levine, R., Liu, G., Liu, G., Locke, K., Macdonald, P., Margulis, N., McGarrity, M., McMan, P., McGurk, A., McKernan, K., McPeckers, R., Melnick, J., Meneus, L., Mihova, T., Miranda, C., Mlenga, V., Morrow, J., Murphy, J., Naylor, J., Norman, C. H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, T. M., Oliver, J., Peterson, K., Pierre, N., Pisan, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D., Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B., Stange, Thoman, N., Stojanovic, N., Subramanian, A., Talamas, J., Testa, S., Theodore, J., Tirrell, A., Travers, M., Trifillo, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G., Zainoun, J., Zimmer, A. and Zody, M.
Submitted (24-APR-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA (bases 1 to 189768)			Levine, R., Liu, G., Liu, G., Locke, K., Lander, E., Abraham, H., Allen, N., Anderson, S., Baldwin, J., Barra, N., Bastien, V., Beda, F., Boguslavsky, L., Boukhalter, B., Brown, A., Burkett, G., Campiano, A., Castle, A., Choepel, Y., Colangelo, M., Collins, S., Collamore, A., Cooke, P., DeCarliano, K., Dewar, K., Diaz, J. S., Dodge, S., Donato, M., Doyle, M., Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L., Grand-Pierre, N., Grant, G., Hages, B., Heatford, A., Horton, L., Howland, J. C., Iliev, I., Johnson, R., Jones, C., Kann, L., Kartas, A., Klein, J., Labocque, K., Lamazares, R., Landers, T., Lechoczy, J., Levine, R., Liu, G., Liu, G., Locke, K., Macdonald, P., Margulis, N., McGarrity, M., McMan, P., McGurk, A., McKernan, K., McPeckers, R., Melnick, J., Meneus, L., Mihova, T., Miranda, C., Mlenga, V., Morrow, J., Murphy, T., Naylor, J., Norman, C. H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, T. M., Oliver, J., Peterson, K., Pierre, N., Pisan, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D., Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B., Stange, Thoman, N., Stojanovic, N., Subramanian, A., Talamas, J., Testa, S., Theodore, J., Tirrell, A., Travers, M., Trifillo, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G., Zainoun, J., Zimmer, A. and Zody, M.
Submitted (24-APR-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA On May 22, 2000 this sequence version replaced gl:7543856. All repeats were identified using RepeatMasker: Smit, A.F.A. & Green, P. (1996-1997) http://ftp.genome.washington.edu/RM/RepeatMasker.html			
Center: Whitehead Institute/ MIT Center for Genome Research Center code: W1BR Web site: http://www-seq.wi.mit.edu Contact: sequence_submissions@genome.wi.mit.edu Project Information Center project name: L9594 Center clone name: 792.C.1			
Sequencing vector: MJ3; M7815; 100% of reads Chemistry: Dye-terminator Big Dye; 100% of reads Assembly program: Phrap; version 0.960731 Consensus quality: 168749 bases at least Q40 Consensus quality: 180047 bases at least Q20 Consensus quality: 184014 bases at least Q20 Insert size: 190000; agarose-fp Insert size: 166268; sum-of-coverage Quality coverage: 3.8 in Q20 bases; agarose-fp Quality coverage: 3.9 in Q20 bases; sum-of-coverage			
* NOTE: This is a 'working draft' sequence. It currently * consists of 36 contigs. The true order of the pieces * is not known and their order in this sequence record is * arbitrary. Gaps between the contigs are represented as * runs of N, but the exact sizes of the gaps are unknown. * This record will be updated with the finished sequence * as soon as it is available and the accession number will * be preserved. * * 1 1015: contig of 1015 bp in length * 1016 1115: gap of 100 bp			


```

Db 86133 GAGAGCCCAAGCTTGGACCTTCACATGTGTGTTCAGTCAAGTACACAGAGACTCA 86074
Qy 263 ArgllleGlValTTPGluProIleArgPheGlnLyslleTyrGlyLysProTrrMet 282
Db 86073 CAAATTGGGGTTTGGGAACTCCACCTAGANTGACATGATGATGAATAATGCCCGGATG 86014
Qy 283 ProArgGlnLysPheAlaValGlyValGlySerSerTrrPrrGhrSerAlaArgVal 302
Db 86013 ACCAGGCAAGAACTTGGCGAGGGGCGGCTTCATGGAACCTTCCTAGGGCAGAG 85954
Qy 303 GlnLysGlyAsnValGlyTTPGluProIleHisArgValProSerGlyAlaProSerSer 322
Db 85953 CAGAAAGAAATATGTAGGTTCAGAGCCACACACAGAGTACCCTAGAGCAGCTTACT 85894
Qy 323 ArgAlaValArgArgSerProProSerSerArgLysGlnLysGlyArgSerThrAspSer 342
Db 85893 GGAGCTGTAGAGAAAGGCCACCATCTCCAGACTCAGAAATGTAGTCCACAAACAGC 85834
Qy 343 LeuGlnHisValProGluLysSerThrAspThrGlnCysGlnProValLysAlaAlaGly 362
Db 85833 TTGCACCTGTGACCTGAGAAAGCCACAGGCACTCATTTGCCAGCTGTGAAGAGCCAG 85774
Qy 363 MetGluSerValProTyrLysThrValValAlaGluLeuThrLysThrValGlylleTyr 382
Db 85773 AGGAGGCTATCCCTGCAGAAAGTACAGAGGGCGAGAGCTGTCCAAAGCTTTGGAGACTAC 85714
Qy 383 LeuLeuHisCysHisAspLeuAspValArgHisGlyValLysArgAspHisPheGlyAla 402
Db 85713 CTTCTGGCTTCAGACCTGACCCGATGTGAGACATGAGACTCAAGAGACATATTTTGAGCT 85654
Qy 403 LeuArgPheAspCysProThrGlyPheArgThrTyrMetGlyProValProLeuCysPhe 422
Db 85653 TTAAAGATTGATGCTGCTGCTGATTCAGACTGCATGCGGCGCTGTAGCCCTTTGCTTT 85594
Qy 423 GlyGlnPhePheProPheGlyTThrAlaValPheThrGlnCysLeuTyrLeuHisCys 441
Db 85593 GGCCAATTTCCTCCATTGGAAACAGCTGATTTATCCCATGCTGTATCCCATTTGT 85537

RESULT 7
AC106901 91323 bp DNA linear PRI 03-APR-2002
LOCUS Homo sapiens BAC clone RP11-1178E6 from 2, complete sequence.
DEFINITION AC106901
ACCESSION AC106901.4 GI:19570233
VERSION HTG.
KEYWORDS Homo sapiens.
SOURCE Homo sapiens.
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

REFERENCE
AUTHORS 1 (bases 1 to 91323)
TITLES Swaenigen-Shahid, S., Kozlowicz, A. and Dignan, G.
JOURNAL Toward a complete human genome sequence
UNPUBLISHED (2001)
GENOME RES. 8 (11), 1097-1108 (1998)
JOURNAL 99063792
PUBMED 9847074

REFERENCE
AUTHORS 2 (bases 1 to 91323)
TITLES Swaenigen-Shahid, S., Kozlowicz, A. and Dignan, G.
JOURNAL The sequence of Homo sapiens BAC clone RP11-1178E6
UNPUBLISHED (2001)
GENOME RES. 8 (11), 1097-1108 (1998)
JOURNAL 99063792
PUBMED 9847074

REFERENCE
AUTHORS 3 (bases 1 to 91323)
TITLES Swaenigen-Shahid, S., Kozlowicz, A. and Dignan, G.
JOURNAL The sequence of Homo sapiens BAC clone RP11-1178E6
UNPUBLISHED (2001)
GENOME RES. 8 (11), 1097-1108 (1998)
JOURNAL 99063792
PUBMED 9847074

REFERENCE
AUTHORS 4 (bases 1 to 91323)
TITLES Swaenigen-Shahid, S., Kozlowicz, A. and Dignan, G.
JOURNAL The sequence of Homo sapiens BAC clone RP11-1178E6
UNPUBLISHED (2001)
GENOME RES. 8 (11), 1097-1108 (1998)
JOURNAL 99063792
PUBMED 9847074

REFERENCE
AUTHORS 5 (bases 1 to 91323)
TITLES Swaenigen-Shahid, S., Kozlowicz, A. and Dignan, G.
JOURNAL The sequence of Homo sapiens BAC clone RP11-1178E6
UNPUBLISHED (2001)
GENOME RES. 8 (11), 1097-1108 (1998)
JOURNAL 99063792
PUBMED 9847074

```

```

AUTHORS Waterston, R. H.
TITLE Direct Submission
JOURNAL Submitted (21-MAR-2002) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
6 (bases 1 to 91323)
AUTHORS Waterston, R.
TITLE Direct Submission
JOURNAL Submitted (03-APR-2002) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
On Mar 21, 2002 this sequence version replaced gi:19424673.
COMMENT
----- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc
Contact: saplens@wustl.wustl.edu
----- Summary Statistics
Center project name: H_NH1178E06
-----

NOTICE: This sequence may not represent the entire insert of this
clone. It may be shorter because we only sequence overlapping
clone sections once, or longer because we provide a small overlap
between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded, sequenced with an alternate
chemistry, or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by sequence
from more than one subclone; and the assembly was confirmed by
restriction digest.

MAPPING INFORMATION:
Mapping Information for this clone was provided by Dr. John D.
McPherson, Department of Genetics, Washington University, St. Louis
MO. For additional information about the map position of this
sequence, see http://genome.wustl.edu/gsc

SDURCE INFORMATION:
The RP11-11 human BAC library was made from the blood of one male
donor, as described by Osoegawa, K., Moon, P.Y., Zhao, B., Frengen, E.,
Tateno, M., Catanese, J.J. and de Jong, P.J. (1998) An improved
approach for construction of bacterial artificial chromosome
libraries. Genomics 51:1-8. The clone may be obtained either from
Research Genetics, Inc. (http://www.resgen.com) or Pieter de Jong
and coworkers at http://www.chori.org
VECTOR: pBAC3.6

NEIGHBORING SEQUENCE INFORMATION:
The clone sequenced to the left is RP11-507C18, 2000 bp overlap;
the clone sequenced to the right is RP11-259I20, 2000 bp overlap.

AC106901 contains a transposon in the growth of the vector.
FEATURES
source
location/Qualifiers
1..91323
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="2"
/map="2"
/clone="RP11-1178E6"
/clone_lib="RP11-11"
28..125
/rpt_family="L1"
126..435
/rpt_family="L1"
436..589
/rpt_family="L1"
590..821
/rpt_family="L1"
822..913
/rpt_family="L1"
914..1074
/rpt_family="L1"
1075..1618
/rpt_family="L1"
1619..1618

```

```
repeat_region /rpt_family="Alu"
1704. .1759
repeat_region /rpt_family="GA-rich"
1949. .2123
repeat_region /rpt_family="L1"
2652. .3582
repeat_region /rpt_family="L1"
3583. .4287
repeat_region /rpt_family="ERV1"
4295. .5073
repeat_region /rpt_family="L1"
5074. .5105
repeat_region /rpt_family="(CAAA)n"
5106. .5232
repeat_region /rpt_family="L1"
5239. .5956
repeat_region /rpt_family="L1"
8153. .8284
repeat_region /rpt_family="L1"
8335. .8526
repeat_region /rpt_family="L1"
8527. .8644
repeat_region /rpt_family="MaLR"
8633. .8771
repeat_region /rpt_family="MaLR"
8781. .10442
repeat_region /rpt_family="L1"
11035. .11325
repeat_region /rpt_family="Alu"
12515. .12689
repeat_region /rpt_family="MER1_type"
13044. .13144
repeat_region /rpt_family="MIR"
13264. .13307
repeat_region /rpt_family="(CA)n"
13835. .14528
repeat_region /rpt_family="ERVK"
15108. .15333
repeat_region /rpt_family="ERV1"
15588. .15692
repeat_region /rpt_family="MIR"
17652. .17965
repeat_region /rpt_family="Alu"
17966. .18124
repeat_region /rpt_family="(TA)n"
18125. .18164
repeat_region /rpt_family="AT-rich"
18350. .18387
repeat_region /rpt_family="MER1_type"
18388. .18659
repeat_region /rpt_family="Alu"
18660. .18746
repeat_region /rpt_family="MER1_type"
19155. .19257
repeat_region /rpt_family="Achobo"
19857. .20350
repeat_region /rpt_family="MaLR"
20354. .20447
repeat_region /rpt_family="MIR"
20477. .20680
repeat_region /rpt_family="MIR"
21333. .21520
repeat_region /rpt_family="L1"
22232. .22307
repeat_region /rpt_family="(TAGA)n"
22506. .22720
repeat_region /rpt_family="MER1_type"
22738. .23033
repeat_region /rpt_family="ERV1"
23041. .23231
repeat_region /rpt_family="ERV1"
23330. .23413
repeat_region /rpt_family="MaLR"
```

```
repeat_region 23474. .23666
/rpt_family="MaLR"
repeat_region 23789. .23810
/rpt_family="AT-rich"
repeat_region 24350. .24427
/rpt_family="MER1_type"
repeat_region 25035. .25129
/rpt_family="MIR"
repeat_region 26055. .26135
/rpt_family="L2"
repeat_region 26545. .26767
/rpt_family="MIR"
repeat_region 31261. .31391
/rpt_family="Mariner"
repeat_region 31511. .31567
/rpt_family="CR1"
repeat_region 32245. .32294
/rpt_family="AT-rich"
repeat_region 32743. .32881
/rpt_family="Alu"
repeat_region 33475. .33986
/rpt_family="MER1_type"
```

Alignment Scores:

Pred. No.:	1.12e-55	Length:	91323
Score:	1209.00	Matches:	238
Percent Similarity:	75.80%	Conservative:	22
Best Local Similarity:	69.39%	Mismatches:	69
Query Match:	50.04%	Indels:	14
DB:	9	Gaps:	2

US-09-997-610-2_COPY_18_459 (1-442) x AC106901 (1-91323)

```
QY 114 GlnGluIleSer-LysGlnGlnSerIleGlnGluValThrTrpValLeuLysAlaPh 133
|||||
Db 72465 GAAGAAATTTCAAAAGACGAGCAAAACATTCAAAATTTGATTTGGGCTGTTAAAGCAT 72524
QY 133 eSerPheIleArgGluValGlnHisLysSerSerGlnAsnLeuHisProAspAsnValI 153
|||||
Db 72525 CAGTTTATTAAGGAACAGCAGCATTAAGGTTGAAATTTGACGCTGCAATGCAAT 72584
QY 153 eLysLysLysAsnProPheSerGlnGlnLysPheLysLeuAlaIleCysIleCy 173
|||||
Db 72585 AGAAAAAATATCCCATTTTCTGAGAGAAATTTCAACGCACTACAGCATTTGCTATAG 72644
QY 173 sAsnGlnGlnLeuAsnValAsnProGlnAspAsnGlyGlnAsnIleSerTrpTrpCysG 193
|||||
Db 72645 TAATGAGAGACTGAATGTTATTCACAAGACAAATGAGAAATGCTCCAGCGCATGCA 72704
QY 193 nArgSerSerGlnGlnSerIleLysSerLeuAlaTrpArgProArgG---LysTrpPh 212
|||||
Db 72705 AAGCTCTTATGCGCGCTCCCATCATCAGCGCTGAGGCTTGAGGAAAAATGCTT 72764
QY 212 eCysGlyThrGlyProGlySerLeuCyCysValGlnProArgAspLeuValProCysVa 232
|||||
Db 72765 TTGTGGCCCGGCGCCAGGGTCCCATGCTATGCTCAATCTAGGACTTTGTTCCCTGCT 72824
QY 232 lPro-----ValAsnSerAlaValAlaSerG 241
|||||
Db 72825 CCCAGCGCTCCACTTGTGACTAAATGATCAAGGTCAGCGTCAGGCCATGCTTCAGA 72884
QY 241 uGlyAlaSerProLysProTrpGlnLeuProSerGlyValGlnProValGlyAlaLysLy 261
|||||
Db 72885 GGGTACAAAGCCCAAGCCTTGCGAGCTTCCACATGCTTGTCAGCTTGCGGTACAGAA 72944
QY 261 sSerArgIleGluValTrpGluProProIleArgPheGlnLysIleTrpGlyAsnProTr 281
|||||
Db 72945 GTTAAGAAATGAGGTTTGGGAACCTTGCTGCTTAATTTACAGAAATGATAGAAACACCTG 73004
QY 281 pMetProArgGlnLysPheAlaValGlyValGlySerSerTrpArgThrSerAlaArgVa 301
|||||
Db 73005 GATGTCAGGACGAGCGTTTGTGTCGACGAGGAGGCGCTCATGAGAACCTGCTAGGGC 73064
```

QY 301 ValGlnLysGlyAsnValGlyTyrGlnProHisArgValProSerGlyAlaProSe 321
Db 73065 AGTGGAGATGGAATGTGGTTGAGACCCCAACAGAGTACTGTGGGTACTGCC 73124
QY 321 rSerArgAlaValArgArgSerProProSerSerArgLeuGlnLysGlyArgSerThrAs 341
Db 73125 TACTGGAGCGTGAGAGATGAGCCACCATCTTCCGAGACTCCAGAAATGATGATCCACTGA 73184
QY 341 pSerLeuGlnHisValProGlnLysSerThrAspThrGlnCysGlnProValLysAlaAl 361
Db 73185 CACCTTGCCCCAGTGCCTGGAAAAAGCCAAACAGTCAATGTCAGCCCAATGAAGACAGC 73244
QY 361 agLysMetGlnSerValProTyrLysThrValAlaAlaGlnLeuThrLysThrValGlyAl 381
Db 73245 TGGGAGGAGGCGGTGACCTGCAAAAGCCACAGGGGCAACACTGCCCAAGCAATGGGAC 73304
QY 381 eTyrLeuLeuHisCysHisAspLeuAspValArgHisGlyValLysArgAspHisPheG 401
Db 73305 CCACCTCTTGACATCAGACATGACCTGATGTGAGACATGAGTCAAAAGACATCATTTTGG 73364
QY 401 yAlaLeuArgPheAspCysProThrGlyPheArgThrTyrMetGlyProVal-ProLeu 421
Db 73365 AGCTTAAAGATTGTGACTGCGCTGCTGAGATTGACACTTGATGGGCGCTGACCGCTTT 73424
QY 421 ySPheGlyGlnPhePhePhePheGlyThrAlaValPheThrGlnCysLeuTyrLeuHis 441
Db 73425 GTTTTGGCCAAATCTCTCCATTGTGAAATGCTGTACTTACCAAAATGCTGTACCCCAT 73484
QY 441 ySmet 442
Db 73485 GTATC 73489
RESULT 8
AL139216 112515 bp DNA linear PRI 25-OCT-2001
LOCUS AL139216
DEFINITION Human DNA sequence from clone RP5-1017J17 on chromosome
ACCESSION AL139216
VERSION AL139216 GI:16501095
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
AUTHORS Ellington, A.
TITLE Direct Submission
JOURNAL Submitted (25-OCT-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: hunquerry@sanger.ac.uk
COMMENT
On Oct 26, 2001 this sequence version replaced gi:15020340.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submissions
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest. The following
abbreviations are used to associate primary accession numbers given
in the feature table with their source databases: Em: EMBL, Sw: SWISSPROT, Tr: TREMBL, Wp: WORMPEP, information on the WORMPEP
database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormep
This sequence was generated from part of bacterial clone compgs of human
chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping
Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr1

RP5-1017J17 is from the library RPT-5 constructed by the group of
Pleier de Jong. For further details see
http://www.chori.org/bacpac/home.htm
VECTOR: PCYPAC2
This sequence is the entire insert of clone RP5-1017J17 The true
left end of clone RP5-944N15 is at 108725 in this sequence.
FEATURES
source
1..112515
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="1"
/map="p31.2-32.2"
/clone="RP5-1017J17"
/clone_11b="RPT-5"
69376..70016
/note="CpG island"
/evidence="not_experimental"
74591..75390
/note="CpG island"
/evidence="not_experimental"
misc_feature
misc_feature
BASE COUNT 33418 a 20758 c 21055 g 37284 t
ORIGIN
Alignment Scores:
Pred. NO.: 1.83e-55 Length: 112515
Score: 1207.00 Matches: 242
Percent Similarity: 66.67% Conservative: 34
Best Local Similarity: 58.45% Mismatches: 83
Query Match: 49.96% Indels: 55
DB: 9 Gaps: 4
US-09-997-610-2_copy_18_459 (1-442) x AL139216 (1-112515)
QY 42 SerAlaPheThrValLysLeuSerGlyLysLeuProLeuProPheLysProLeuPhe 61
Db 25088 TCTTGTTTACCAAAAAGAGTGTGACATTTTGCTGCTGCTTACAGATTGCTG- 25141
QY 62 ThrGlyValLeuTyrAsnAlaGlnArgAspLeuLysGluAlaMetGlyValPheAlaCys 81
Db 25142 -----GAACTTGAACCTTGAGAGAGATTAGGCTATCTGGCA--- 25177
QY 82 ArgValProGlnsnnTyrTyrSerSerPheAspValGlnLeuHisCysLysValAsn 101
Db 25177 ----- 25177
QY 102 IleThrLeuMetArgLysGlnIleLeuAlaAsnLysGlnGluLysSerGlnGlnSer 121
Db 25178 -----GAAAGAAATTTCTAACCAACGAAGC 25201
QY 122 IleGlnGluValThrTrpValLeuLeuLysAlaPheSerPheIleArgGluAlaGlnHis 141
Db 25202 ATTCAGAGGTGACTGGGTGGTGTAAAGCATTCAGTTTATTAAGGAGAGACGAT 25261
QY 142 LysSerSerLysLysLeuHisProAspAsnValIleLysLysLysAsnProPheSerGlu 161
Db 25262 AAAAGATTGGAAAAATTTTACCTGACATTCGAATTAAGAAAAATATCCATTTTGTGAA 25321
QY 162 GlyLysPheLysLeuAlaAlaGlnIleCysIleCysAsnGlnGluLeuAsnValAsnPro 181
Db 25322 ATGAATTTCAAGCTGGCTGCAGAAATGTGATGAATGAGAACCAAAATTTAATCAC 25381
QY 182 GlnAspAsnGlyLysAsnIleSerTrpThrCysGlnArgSerSerGlnGlnSerIleLys 201
Db 25382 CAAGACAAATGGGGAATAATGTCACAGACATGTGAGAGTCTTCATGSCAGCCCTCGCA 25441
QY 202 SerLeuAlaTrpArgProArgArg---LysTrpPheCysGlyThrGlyProGlySerLeu 220
Db 25442 TCACAAAGCTGTGAGAGCTGTGAGAGTAATGTGTTTGTGGAGTGGGCGCCAGGTTCCCGG 25501
QY 221 CysCysValGlnProArgAspLeuValProCysValProValAsnSerAlaVal----- 238
Db 25502 TGTGTGTGAGTGTAGTATTGTTGTGGCTGCAATCCAGGCAAGCTCCAGCGGTGACTAAA 25561

```

Qy 239 -----AlaSerGluGlyAlaSerProLysProTyrGln 249
Db 25562 AGGATGAAGAGTACAGCTCAAGCTATGGCTGAGAGGCTGAAGCCCAAGCCCTGGCAG 25621
Qy 250 LeuProSerGlyValGluProValGlyAlaLysSerArgGluValTTPGluPro 269
Db 25622 CTTCCATGTCGTGAGCTGTGAGTGCACAGATGTCAGAACTGAGTATGAGAACCT 25681
Qy 270 ProILArpGheGlnLysIleTyrGlyAsnProTyrMetProArgGlnLysPheAlaVal 289
Db 25682 CCCCCTAGATTTTGAAGAGATGATGATGACGCTGATGCCAGCCAGAAAGTTTGCAGCA 25741
Qy 290 GlyValGlySerSerTyrPargTnrsSerAlaArgValGluGlnLysGlyAsnValGlyTyr 309
Db 25742 GGGGCGAGGGCCATCAGAGAAAGCTGCTGACAGGCAATGCAAGAAAGAAATGGGGTGG 25801
Qy 310 GluProProHlaArgValProSerGlyAlaProSerSerArgAlaValArgArgSerPro 329
Db 25802 GAGCCCCCAACACAGAGTCCCTACTGGGGGCACTGCTGAGTGGAGCTTGAGAGAGACCA 25861
Qy 330 ProSerSerArgLeuGlnLysGlyArgSerThrAspSerLeuGlnHisValProGluLys 349
Db 25862 CCATCTCCAGACCCCAAGATGATGATCCAGCTGACAGCTTGACACCTGTCAGGAGAAA 25921
Qy 350 SerThrAspThrGlnCysGlnProValAlaAlaGlyMetGluSerValProTyrLys 369
Db 25922 GCCATGACACCTCAGCCCAAGATGATGATCCAGCTGACAGCTTGACACCTGTCAGGAGAAA 25981
Qy 370 ThrValValAlaGluLeuThrLysThrValGlyTyrLeuLeuHisCysHisAspLeu 389
Db 25982 GCCACAGAGGGGTGAGCTGCTCCAGAGACCATGCCACTTGCATCAGTGTGACCCG 26041
Qy 390 AspValArgHisGlyValLysArgAspHisPheGlyAlaLeuArgPheArgProThr 409
Db 26042 GATGCAAGATACGAGAGTGAAGAAAGATCATTTTGAAGTTTAAATTTGACTGCCCTGCT 26101
Qy 410 GlyPheArgThrTyrMetGlyProVal-ProLeuCysPheGlyGlnPhePheProPheG1 429
Db 26102 GGAATTTGGACCTGCATGGGGGCTGTAGGCCCTTTGTTGGCCAAATTTCTCCATTGG 26161
Qy 429 yThrAlaValPheThrGlnCysLeuTyrLeuHisCysMet 442
Db 26162 ATCAACTGTATTTATCCCAATGCTGTAGCCCATTTGATC 26201

RESULT 9
LOCUS HS345P10 137785 bp DNA linear PRI 12-DEC-1999
DEFINITION Human DNA sequence from clone RP3-345P10 on chromosome 22. Contains
a pseudogene similar to 14-3-3 protein theta-subtype, parts of 2
novel genes based on ESTs, a testis-specific mRNA, ESTs, STSs and
GSSs, complete sequence.
ACCESSION Z82201
VERSION Z82201.1 GI:1843447
KEYWORDS HTG: 14-3-3 protein.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 137785)
AUTHORS Whiteley, M.
TITLE Direct Submission
JOURNAL Submitted (08-DEC-1999) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk; Clone
requests: clonerequest@sanger.ac.uk
COMMENT
On Feb 19, 1997 this sequence version replaced gi:1772951.
This sequence has been finished according to sequence map criteria
as follows. An attempt is made to resolve all sequencing problems,
such as compressions and repeats, but not necessarily within known
annotated human repeat sequence elements (e.g. Alu). Where the
sequence is ambiguous, there is an annotation using the 'unsure'
feature key.
This sequence was generated from part of bacterial clone contigs of
human chromosome 22, constructed by the Sanger Centre Chromosome 22

```

FEATURES

```

source
1..137785
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="22"
/clone="RP3-345P10"
/clone_lib="RP3-3"
194..497
repeat_region
/note="AluSP repeat: matches 1..308 of consensus"
1583..1748
repeat_region
/note="LINE repeat: matches 5464..5634 of consensus"
1936..2241
repeat_region
/note="AluSX repeat: matches 10..298 of consensus"
2451..2689
repeat_region
/note="MIR repeat: matches 10..262 of consensus"
3104..3436
misc_feature
/note="match: GSS: Em:AQ702221"
4276..4346
repeat_region
/note="MIR repeat: matches 97..181 of consensus"
4886..4995
repeat_region
/note="MIR repeat: matches 59..175 of consensus"
5352..6731
repeat_region
/note="69 copies 20 mer 88 conserved"
301n(7248..7402,16817..16920)
/gene="d3345P10.4"
301n(<7248..7402,16817..>16920)
/gene="d3345P10.4"
/product="d3345P10.4 (testis-specific mRNA)"
/note="match: ESTs: Em:AF012350 Em:AF012385"
/evidence="not_experimental"
8997..9586
/gene="d3345P10.4"
/note="match: GSS: Em:AQ344456"
9079..9118
repeat_region
/note="20 copies 2 mer ac 95 conserved"
9137..9384
repeat_region
/note="HAL1 repeat: matches 698..949 of consensus"
9549..10178
repeat_region
/note="MER92B repeat: matches 5..636 of consensus"
10282..10637
repeat_region
/note="MTF1A1 repeat: matches 1..365 of consensus"
10967..13081
repeat_region
/note="LIMCB repeat: matches 196..2309 of consensus"
13084..13197
repeat_region
/note="AluJo/FRAM repeat: matches 195..308 of consensus"
13466..13641
repeat_region
/note="MER5A repeat: matches 1..189 of consensus"
14078..14344
repeat_region
/note="MTF1D repeat: matches 1..501 of consensus"
14674..14815
repeat_region
/note="MER53 repeat: matches 27..186 of consensus"

```

Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/Chr22>

During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em, EMBL; Sw, SWISSPROT; Tr, TrEMBL; Wp, WormPEP; Information on the WormPEP database can be found at <http://www.sanger.ac.uk/projects/C.elegans/wormpep>

RP3-345P10 is from the library RPCI-3 constructed at the Roswell Park Cancer Institute by the group of Pieter de Jong. For further details see <http://bacpac.med.buffalo.edu/>

VECTOR: pCYPAC2

IMPORTANT: This sequence is not the entire insert of clone RP3-345P10. It may be shorter because we only sequence overlapping sections once, or longer because we arrange for a small overlap between neighbouring submissions.

The start of this sequence overlaps with sequence AL023801. Location/Qualifiers

```

1..137785
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="22"
/clone="RP3-345P10"
/clone_lib="RP3-3"
194..497
repeat_region
/note="AluSP repeat: matches 1..308 of consensus"
1583..1748
repeat_region
/note="LINE repeat: matches 5464..5634 of consensus"
1936..2241
repeat_region
/note="AluSX repeat: matches 10..298 of consensus"
2451..2689
repeat_region
/note="MIR repeat: matches 10..262 of consensus"
3104..3436
misc_feature
/note="match: GSS: Em:AQ702221"
4276..4346
repeat_region
/note="MIR repeat: matches 97..181 of consensus"
4886..4995
repeat_region
/note="MIR repeat: matches 59..175 of consensus"
5352..6731
repeat_region
/note="69 copies 20 mer 88 conserved"
301n(7248..7402,16817..16920)
/gene="d3345P10.4"
301n(<7248..7402,16817..>16920)
/gene="d3345P10.4"
/product="d3345P10.4 (testis-specific mRNA)"
/note="match: ESTs: Em:AF012350 Em:AF012385"
/evidence="not_experimental"
8997..9586
/gene="d3345P10.4"
/note="match: GSS: Em:AQ344456"
9079..9118
repeat_region
/note="20 copies 2 mer ac 95 conserved"
9137..9384
repeat_region
/note="HAL1 repeat: matches 698..949 of consensus"
9549..10178
repeat_region
/note="MER92B repeat: matches 5..636 of consensus"
10282..10637
repeat_region
/note="MTF1A1 repeat: matches 1..365 of consensus"
10967..13081
repeat_region
/note="LIMCB repeat: matches 196..2309 of consensus"
13084..13197
repeat_region
/note="AluJo/FRAM repeat: matches 195..308 of consensus"
13466..13641
repeat_region
/note="MER5A repeat: matches 1..189 of consensus"
14078..14344
repeat_region
/note="MTF1D repeat: matches 1..501 of consensus"
14674..14815
repeat_region
/note="MER53 repeat: matches 27..186 of consensus"

```

repeat_region	15110. .15434	/note="HAI1 repeat: matches 155. .499 of consensus"
repeat_region	15448. .15632	/note="HAI1 repeat: matches 1094. .1296 of consensus"
repeat_region	15772. .15873	/note="HAI1 repeat: matches 1431. .1533 of consensus"
repeat_region	15878. .15988	/note="AluY repeat: matches 1. .111 of consensus"
repeat_region	15997. .16043	/note="Alu repeat: matches 255. .301 of consensus"
repeat_region	16044. .16067	/note="12 copies 2 mer aa 95 conserved"
repeat_region	16249. .16563	/note="AluX repeat: matches 1. .309 of consensus"
repeat_region	17306. .17475	/note="L2 repeat: matches 2527. .2710 of consensus"
repeat_region	17821. .17886	/note="33 copies 2 mer tt 81 conserved"
repeat_region	17832. .17891	/note="3 copies 20 mer 85 conserved"
repeat_region	17919. .18229	/note="AluS9 repeat: matches 1. .310 of consensus"
misc_feature	19712. .20092	/note="match: GSS: Em:AQ166214"
repeat_region	20737. .21046	/note="AluSP repeat: matches 1. .311 of consensus"
repeat_region	21079. .24772	/note="LIM4 repeat: matches 2179. .5810 of consensus"
repeat_region	24773. .25081	/note="AluJb repeat: matches 1. .309 of consensus"
repeat_region	25082. .25525	/note="LIM4 repeat: matches 5810. .6286 of consensus"
repeat_region	25545. .25705	/note="LIM5 repeat: matches 6128. .6295 of consensus"
repeat_region	26004. .26282	/note="AluX repeat: matches 34. .312 of consensus"
repeat_region	26410. .26509	/note="L2 repeat: matches 2583. .2709 of consensus"
gene		complement(join(26812. .27006,35225. .35343,42569. .42665))
CDS		complement(join(266812. .27006,35225. .35343,42569. .>42665)))
	/gene="d3345P10.2"	/note="match: ESTs: Em:AA497915"
	/codon_start=2	/evidence=not_experimental
	/product="d3345P10.2 (part of a novel gene)"	/protein_id="CAB63029.1"
	/db_xref="GI:5572210"	/translation="EKQVAVSEIACEDAHQVLYTAKNRKSDLSKNELETNDGNGI
	LRRRRIRKMLYGFDIRPIPRREFEKLARVYDEKSGHTYDFEFLKGINYSFAVHRPC	AEDYFNFMGHGFKPQOLDEMKELQSTERAFAA"
repeat_region	27124. .27206	/note="MER5A repeat: matches 9. .93 of consensus"
repeat_region	27395. .27705	/note="AluY repeat: matches 1. .311 of consensus"
repeat_region	30575. .30981	/note="MST4 repeat: matches 1. .426 of consensus"
misc_feature		complement(31305. .31756)
	/gene="d3345P10.2"	/note="match: GSS: Em:AQ138484"
	/complement(31307. .31543)	/gene="d3345P10.2"
	/note="match: GSS: Em:AQ269272"	31874. .32097
repeat_region	32106. .32395	/note="LTR8 repeat: matches 1. .227 of consensus"
repeat_region	32396. .32544	/note="AluSC repeat: matches 1. .291 of consensus"
repeat_region	33191. .33417	/note="LIM3 repeat: matches 6028. .6181 of consensus"
repeat_region	35858. .35926	/note="MIR repeat: matches 157. .222 of consensus"

[illegible]

Oy	237	-----	-AlaValIAserGluGIYAlaseProlys	246
Db	78023	ATGGCTGAAGAAGGGCCAAATTTAGACCTTGGGCCAATGCTTCAGAGGAGCAACCGTTCAAG		78082
Oy	247	ProTrpGlnLeuProSerGlyValGluProValGlyAlaLysSerArgIleGluVal		266
Db	78083	CCTTGAGCAGCTTCCACAGATGGTGTGAGGCCTGGAGTGCACAAGAACTCAAGAATTGAGGTT		78142
Oy	267	TrpGluProProlIeArgPheGlnLysIleTyrgLysAnpProTrpMetProArgGlnLys		286
Db	78143	TGGGAACCTCTGCCAGATTCTCAGAAGATCTCACGGAAAACGCTCGATGCTCAAGCAGAAAG		78202
Oy	287	Pheala-ValGlyValGlySerSerTrpArgTrpSerIlaArguaValGlnLysGlyAs		306
Db	78203	TTTTGCCCCGAGGGGAGGGCTCTCTATACACAACCTCTGCTGAGGCAAGTGCMAAAGGANA		78262
Oy	306	nValGlyTrpGluProProlHisArgValIProSerGlyLaIProSerSerArgAlaValAr		326
Db	78263	TGTGGGGGACAGGGCCCCACACAGAGTCCCTACTGGGGCACTGCCATGGAGAGCTGTAG		78322
Oy	326	GArgSerProProSerSerSerArgLengLnLysGlyArgSerThrAspSerLeuGlnHisVa		346
Db	78323	AAGAGGGCCACCATCTCTCAAACCACACAGCGTGAATCCACTGAAGCTTGACCAATGC		78382
Oy	346	IProGlyLysSerThrApTrhngLnCsGlnProValLysAlaAlaGlyMetGlnSerVa		366
Db	78383	TCTCGAAAAAGCTGACAGCACTCAATGTGCAGCCCACGAAAGCAGCGGAGGAGGAGCTGT		78442
Oy	366	IProGlyLysThrValValAlaGluLeuThrLysThrValGlyIleTyrgLeuGlnHisCy		386
Db	78443	ACCCTGAACCAACACAGGGGGGAGACTCCCAAGAACCATGGGAACCACTGTACATCA		78502
Oy	386	SHiasPreuAspValArghHisGlyValLysArgAspHisSperGlyAlaLeuArPheas		406
Db	78503	GTGTGACCTGGATGTGACATGAGTAAGAAGATCATTTTAGAGCTTTAAAGATTGGA		78562
Oy	406	pCyseroThrcylgPheaqThrTyrmcGlYProVal-I-ProleucycspheglugInPheP		426
Db	78563	CTGCCCCACCTGATTTTCAGACTGCATGGGCGCTTACGCCATTTGTTTGGCCAAATTTC		78622
Oy	426	heProPheGlyThrAlaValIpherhrgLnCsLeuTyrgLeuHISCySmet		442
Db	78623	TCCCATTTTGAATGCGCTGATTATACCACAACCTGACCTCATCTGTATATC		78672
RESULT 10	AC009047/c			
LOCUS	AC009047	168080 bp	DNA linear	HTG 21-JUN-2000
DEFINITION	Homo sapiens chromosome 16 clone RP11-204G18, WORKING DRAFT			
SEQUENCE	37 unordered pieces.			
ACCESSION	AC009047			
VERSION	AC009047.5	GI:8575942		
KEYWORDS	HTG; HTGS_PHASE1; HTGS_DRAFT.			
SOURCE	Homo sapiens.			
ORGANISM	Homo sapiens			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
AUTHORS	Mammalia; Euthelia; Primates; Catarrhini; Homnidae; Homo.			
JOURNAL	DOE Joint Genome Institute.			
REFERENCE	Sequencing of Human Chromosome 16			
AUTHORS	Unpublished			
JOURNAL	2 (bases 1 to 168080)			
REFERENCE	DOE Joint Genome Institute.			
AUTHORS	Direct Submission			
JOURNAL	Submitted (03-AUG-1999) Production Sequencing Facility, DOE Joint			
COMMENT	Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA			
	On Jun 21, 2000 this sequence version replaced gi:7689877.			
	-----Genome Center			
	Center: Joint Genome Institute			
	Center Code: JGI			
	Web site: http://www.jgi.doe.gov			

	Project Information			

```

Center Project Name: 148863
Center clone name: RPCT_11_204G18
-----
Summary Statistics
Consensus quality: 140999 bases at least Q40
Consensus quality: 155370 bases at least Q30
Consensus quality: 158861 bases at least Q20
Estimated insert size: 161080; agarose-IP estimation
Estimated insert size: 164480; sum-of-contigs estimation
Quality coverage: 3.7 in Q20 bases; agarose-IP estimation
Quality coverage: 3.62 in Q20 bases; sum-of-contigs estimation.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 37 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence.
* as soon as it is available and the accession number will
* be preserved.
1
1028: contig of 1028 bp in length
1029
1128: gap of unknown length
1129
2312: contig of 1184 bp in length
2313
2412: gap of unknown length
2413
3552: contig of 1140 bp in length
3553
3652: gap of unknown length
3653
5214: contig of 1562 bp in length
5215
5314: gap of unknown length
5315
6523: contig of 1209 bp in length
6524
6623: gap of unknown length
6624
7727: contig of 1104 bp in length
7728
7827: gap of unknown length
7828
9027: contig of 1200 bp in length
9028
9127: gap of unknown length
9128
10410: contig of 1283 bp in length
10411
10510: gap of unknown length
10511
12699: contig of 2189 bp in length
12700
12799: gap of unknown length
12800
14910: contig of 2111 bp in length
14911
15010: gap of unknown length
15011
17850: contig of 2840 bp in length
17851
17950: gap of unknown length
17951
20739: contig of 2789 bp in length
20740
20839: gap of unknown length
20840
23334: contig of 2495 bp in length
23335
23434: gap of unknown length
23435
26582: contig of 3148 bp in length
26583
26682: gap of unknown length
26683
30181: contig of 3499 bp in length
30182
30281: gap of unknown length
30282
33044: contig of 2763 bp in length
33045
33144: gap of unknown length
33145
37402: contig of 4258 bp in length
37403
37502: gap of unknown length
37503
40621: contig of 3119 bp in length
40622
40721: gap of unknown length
40722
45438: contig of 4717 bp in length
45439
45538: gap of unknown length
45539
49262: contig of 3724 bp in length
49263
49362: gap of unknown length
49363
52985: contig of 3623 bp in length
52986
53085: gap of unknown length
53086
57167: contig of 4082 bp in length
57168
57267: gap of unknown length
57268
60143: contig of 2876 bp in length
60144
60243: gap of unknown length
60244
64483: contig of 4240 bp in length
64484
64583: gap of unknown length
64584
66813: contig of 4030 bp in length
66814
68713: gap of unknown length
68714
74972: contig of 6259 bp in length
74973
75072: gap of unknown length
75073
76885: contig of 3613 bp in length
76886
78785: gap of unknown length

```

```

* 78786 83681: contig of 4896 bp in length
* 83782 83781: gap of unknown length
* 83782 88293: contig of 5518 bp in length
* 83782 89399: gap of unknown length
* 89300 89399: gap of unknown length
* 89400 94638: contig of 5239 bp in length
* 94639 94738: gap of unknown length
* 94739 100450: contig of 5712 bp in length
* 100451 100550: gap of unknown length
* 100551 106362: contig of 5812 bp in length
* 106363 106462: gap of unknown length
* 106463 112229: contig of 5767 bp in length
* 112230 112329: gap of unknown length
* 112330 119405: contig of 7076 bp in length
* 119406 119505: gap of unknown length
* 119506 129237: contig of 9732 bp in length
* 129238 129337: gap of unknown length
* 129338 143878: contig of 14541 bp in length
* 143879 143978: gap of unknown length
* 143979 168080: contig of 24102 bp in length.

FEATURES
    source          1. 168080
                    /organism="Homo sapiens"
                    /db_xref="taxon:9606"
                    /chromosome="16"
                    /clone="RP11-204G18"
                    /clone_lib="RPCT human BAC library 11"

BASE COUNT      51798 a 30488 c 30183 g 51984 t 3627 others
ORIGIN
Alignment Scores:
Pred. No.:      7e-55      length:      168080
Score:          1200.00    Matches:      235
Percent Similarity: 76.32%  Conservative: 26
Best Local Similarity: 68.71%  Mismatches: 68
Query Match:    49.67%      Indels:      13
DB:             2          Gaps:      2

US-09-997-610-2_COPY_18_459 (1-442) x AC009047 (1-168080)
QY 114 GIUGIUISErLySGInSerLIeGInGUaIvAlThrTrPaIleuLeuIySaIaPhe 133
      |||||
Db 161996 GAAGAATTTCTAAGCACAAGACATTAAAGAGGTGAGTGGCTGCTATTAAAGCATTC 161937
      |||||
QY 134 SerPheIleArGUaIeGInHISLySerSerGIuSnleuHISProSaPheValIle 153
      |||||
Db 161936 AGTTTATTAAGGAAACAGCATGTTGGAAATTTGCACCTGCTGATGTGATA 161877
      |||||
QY 154 LysIlySAAsnProPheSerGIyLySPheLySerIeUaIaGInLIeCyStIeCy 173
      |||||
Db 161876 GAAGAAGAAATCCATTTTCTGAGAGAGAAATCAAACTGCTCAGAAATTTGCATAGT 161817
      |||||
QY 174 AsnGIUGIleuAsnValaSnProGIaSPaSnGIyGIuSnIleSerTrPrHrcysGIn 193
      |||||
Db 161816 AATGAAGAACAATGTAAATTCACAGACATGGGAAATATGCTCCAGGCGAGTCAA 161757
      |||||
QY 194 ArgSerSerGIInSerIleLySerIeUaIaTrPrArProArGArG---LysTrPhe 212
      |||||
Db 161756 AGCTTTACAGAGCCCTCCATCCACAGGCCAGGCTTGAAGGAGAAAGTGTTT 161697
      |||||
QY 213 CysGIyThrGIyProGIySerIeUaIaGInProArGArPheUaIaProCySaVal 232
      |||||
Db 161696 TGTGGGCTGGGACCAAGGCTCCTGTGCTGTGTTCAGCTTACTTGTGGCTGTATC 161637
      |||||
QY 233 ProValaSn-----SerAlaValaIaSerGIu 241
      |||||
Db 161636 CCAAGCTCTCCAGCAGCAAGGACATAGTAGAACTCAGGACCATGGCTTCAGAA 161577
      |||||
QY 242 GlyAlaSerProLySProTrPrGIInleuProSerGIyValaGInUaIaIaLyS 261
      |||||
Db 161576 GGTACACAGCCCAAGACTTGGACATCTACATGATTGAGCTGTGATGACACTGAAG 161517
      |||||
QY 262 SerArGIleGIuValTrPrGIuProIleArPheGIInLySleTrPrGIuProTr 281
      |||||

```

```

Db 161516 TCAAGATTTGGCTTTGGGAATCTCCACCTAGATTTTCAGAGAGATGTATGAATGCTCG 161457
QY 282 MetProArGIInLySPheAlaValaGIySerSerTrPrArGTrSerAlaArGVal 301
      |||||
Db 161456 ATGACAGGAGAAATGTTGCTCAGAGGTGGGCCCTCATAGATATCCCTGCTAGGGCA 161397
      |||||
QY 302 ValGIInLySGInSAValGIyTrPrGIuProProHISArGValaProSerGIyAlaProSer 321
      |||||
Db 161396 GTGTGAAGAGGAAATGTGGTTTGGAACTCCACACAGATCCCTACTGTGGGCCCACT 161337
      |||||
QY 322 SerArGIaValaIaArGArSerProProSerArGIeGIInLySGIyArGSerThraSP 341
      |||||
Db 161336 AGTGGACCTGTGGAAGAGACACCATCTTCACAGACCCCAAGATGGCCATCTACTGAC 161277
      |||||
QY 342 SerLeuGIInHISValProGIuLySerThraSPHrGIInLySGInProValaIaAla 361
      |||||
Db 161276 AGCTTGACCTGTGCTGGAAGAGACACATCATCCAGTCCATCGAAGAAAGAGCT 161217
      |||||
QY 362 GlyMetGIuSerValProTrLySlyThraValaIaGIuLeuThrLySlyThraGIyIle 381
      |||||
Db 161216 GGGAGGAGAGCTGTACCCAGCAGACCCAGGACAGGACAGCTGCTGAGACCATGGAMCC 161157
      |||||
QY 382 TyrLeuLeuHISLySGInSAValaIaArGIHISGIyValaLySArGASPHISPhEGly 401
      |||||
Db 161156 CAGCTTTGATCATGATATACAGACCCAGACATGAGTCAGACATATTTTGA 161097
      |||||
QY 402 AlaLeuArPheArPheSPProThrGIyPheArGIyThrTrMetGIyProVal-ProLeuCy 421
      |||||
Db 161096 GCTTTAAGATTTGACTGCCCCACCTGATTTTCAAGCTTGCAGGAGGCGCAGTGGCCCTTTG 161037
      |||||
QY 421 SPHeGIyInPhePheProPheGIyThraIaIaPheThrGIInLySGInLySleuThra 441
      |||||
Db 161036 TTTTACCAATTTCTTCATTTGAATGCTGTATTACCAATGCTGTATCCCATTTG 160977
      |||||
QY 441 smet 442
      |||||
Db 160976 TATC 160973

RESULT 11
AC093536/ 199579 bp DNA linear PRI 09-JUL-2002
LOCUS Homo sapiens chromosome 16 clone RP11-744D14, complete sequence.
DEFINITION AC093536 AC021186
VERSION AC093536.2 GI:21717113
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 199579)
AUTHORS DOE Joint Genome Institute.
TITLE Sequencing of Human Chromosome 16
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 199579)
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (31-AUG-2001) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
REFERENCE 3 (bases 1 to 199579)
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (09-JUL-2002) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT This entry has been annotated with sequence quality
estimates computed by the Phrap assembly program.
All manually edited bases have been reduced to quality zero.
Quality levels above 40 are expected to have less than
1 error in 10,000 bp.
Base-by-base quality values are not generally visible from the
GenBank flat file format but are available as part
of this entry's ASN.1 file.

```


clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:

Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:

The RPCI-11 human BAC library was made from the blood of one male donor, as described by Osoegawa, K., Woon, P.Y., Zhao, B., Frengen, E., Tateno, M., Catanese, J.J., and de Jong, P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at the Roswell Park Cancer Institute (<http://bacpac.med.buffalo.edu>)

VECTOR: pBACe3.6

NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the left is Rp11-318p5, 2000 bp overlap; the clone sequenced to the right is Rp11-299014, 2000 bp overlap. Actual start of this clone is at base position 153702 of Rp11-318p5; actual end is at base position 4782 of Rp11-299014.

Data from AC016715 was used to finish this clone, AC011238.

Polymorphisms have been identified between AC016715 and AC011238.

Location/Qualifiers

FEATURES

source

```
1. 155764
   /organism="Homo sapiens"
   /db_xref="taxon:9606"
   /chromosome="2"
   /map="2"
   /clone="Rp11-374C5"
   /clone_1id="RPCI-11"
   195..589
   /rpt_family="L2"
repeat_region
374..400
misc_feature
   /note="match to EST A1471723 (NID:g4333813) ta16a05.x1"
628..840
misc_feature
   /note="match to EST AA443394 (NID:g2156069) zw87b01.r1"
628..833
misc_feature
   /note="match to EST AW952412 (NID:g8142094)"
628..833
misc_feature
   /note="match to EST BE856700 (NID:g10369992)"
628..833
misc_feature
   /note="similar to Mus musculus EST BG094183 (NID:g12576746) uu72f01.x1"
628..833
misc_feature
   /note="similar to Rattus norvegicus EST B1289336 (NID:g14946795)"
628..816
misc_feature
   /note="match to EST AA328123 (NID:g1980388)"
633..833
misc_feature
   /note="match to EST A1471723 (NID:g4333813) ta16a05.x1"
645..1151
misc_feature
   /note="match to EST BE4466172 (NID:g9511947) hz60c11.x1"
788..821
misc_feature
   /note="match to EST N39733 (NID:g1163278) yx92g01.r1"
1108..1191
misc_feature
   /note="match to EST R51616 (NID:g813518) y977b04.s1"
1593..1870
repeat_region
   /rpt_family="Alu"
1665..1667
misc_feature
   /note="match to EST AA625701 (NID:g2538088) zu86c07.s1"
```

```
repeat_region
2644..2671
   /rpt_family="(T)n"
2669..3087
misc_feature
   /note="match to EST AL603583 (NID:g15167089)"
3474..4110
misc_feature
   /note="match to EST BF348692 (NID:g11283118)"
3958..4293
misc_feature
   /note="similar to Rattus norvegicus EST A1058342 (NID:g3332119)"
4046..4311
misc_feature
   /note="match to EST AW952412 (NID:g8142094)"
4046..4305
misc_feature
   /note="match to EST A1471723 (NID:g4333813) ta16a05.x1"
4046..4293
misc_feature
   /note="similar to Rattus norvegicus EST B1289336 (NID:g14946795)"
4046..4247
misc_feature
   /note="similar to Mus musculus EST BG094183 (NID:g12576746) uu72f01.x1"
4049..4305
misc_feature
   /note="match to EST AA625701 (NID:g2538088) zu86c07.s1"
4377..4551
repeat_region
   /rpt_family="MER2_type"
4885..5179
repeat_region
   /rpt_family="Alu"
5612..5735
repeat_region
   /rpt_family="Alu"
5865..6033
repeat_region
   /rpt_family="CRL"
6109..6113
misc_feature
   /note="match to EST AV725468 (NID:g10830886)"
6687..6729
repeat_region
   /rpt_family="AT_rich"
6942..7082
misc_feature
   /note="match to EST A1907606 (NID:g6498041)"
7227..7535
repeat_region
   /rpt_family="Alu"
7230..7257
repeat_region
   /rpt_family="(TTTGG)n"
7543..7609
repeat_region
   /rpt_family="MER1_type"
7963..7990
repeat_region
   /rpt_family="AT_rich"
8501..8549
misc_feature
   /note="match to EST N39733 (NID:g1163278) yx92g01.r1"
8961..9280
misc_feature
   /note="match to EST AA883177 (NID:g2994653) am18a07.s1"
8962..9229
misc_feature
   /note="match to EST BF837074 (NID:g12189239)"
9359..9680
repeat_region
   /rpt_family="Alu"
9636..9676
repeat_region
   /rpt_family="AT_rich"
10114..10445
misc_feature
   /note="match to EST W60003 (NID:g1366762) zc92g08.s1"
10382..10435
repeat_region
   /rpt_family="MIR"
10362..10656
misc_feature
   /note="match to EST AA446382 (NID:g2159047) zw58c01.r1"
10678..10980
repeat_region
   /rpt_family="Alu"
10941..11188
misc_feature
   /note="match to EST AV725468 (NID:g10830886)"
10973..11380
misc_feature
   /note="match to EST AA446382 (NID:g2159047) zw58c01.r1"
11112..11453
misc_feature
   /note="match to EST BE856700 (NID:g10369992)"
11112..11425
misc_feature
   /note="match to EST N39733 (NID:g1163278) yx92g01.r1"
11117..11142
misc_feature
   /note="match to EST AA443394 (NID:g2156069) zw87b01.r1"
11155..11792
misc_feature
   /note="match to EST N26638 (NID:g1140986) yx92g01.s1"
```

```

Db 29191 ACCTAGTAACCTGTGAGCAATAGGGCCATCATCTCCAGACTCCAGAAATGGTAGATCCAC 29132
Qy 340 TASPSETLeuGlnHisValProGluLysSerThrAspThrGlnCysGlnProValLysAl 360
Db 29131 TGACAGCTTTGTACCTATGGCGCTGGAAAGAGCTCAGACACTCAACA -CCGCCAGTGAAGT 29073
Qy 360 aAlaGlyMetGlnSerValProTyrLysThrValValAlaGluLeuThrLysThrValG1 380
Db 29072 AGCTGGGAAGGAGGGCTGTACCTCTGCAGAACCCACAGAGGGTGGGCTGCCCAAGACTATGGG 29013
Qy 380 y1LeuTyrLeuLeuHisCysHisAspLeuAspValArgHisGlyValLysAspAspHis 400
Db 29012 AACCTGCTCTTGCATCATCAGCATGACCTGGATGTGAGCATGAGATGAGCAAGAGATCATTT 28953
Qy 400 eGlyValAlaLeuArgPheAspCysProThrGlyPheArgThrTyrMetGlyProVal -ProL 420
Db 28952 TGGAGCTTTAGATTTGACTGTCCACACGCGGTTTCAGACTGCATGGAGGGCGCTGTAGCCCC 28893
Qy 420 eucysPheGlyGlnPhePheProPheGlyThrAlaValPheThrGlnCysLeuTyrLeuH 440
Db 28892 TTTGTTTGGCCAAATTTCTCCATTAGATTGGCTGTATTAACCCAAATGCCCTGACCCCT 28833
Qy 440 1sCysMet 442
Db 28832 ATTGTATC 28825

RESULT 13
AC016715/c
LOCUS AC016715 187898 bp DNA linear HTG-07-JUL-2000
DEFINITION Homo sapiens chromosome 2 clone RP11-311D8, WORKING DRAFT SEQUENCE.
ACCESSION AC016715
VERSION AC016715.4 GI:8572513
KEYWORDS HTG; HTGS; PHASE1; HTGS_DRAFT.
SOURCE Homo sapiens.
ORGANISM Homo sapiens.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
1 (bases 1 to 187898)
Waterston, R.H.
The sequence of Homo sapiens clone
Unpublished
2 (bases 1 to 187898)
Waterston, R.H.
Direct Submission
Submitted (04-DEC-1999) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
On Jun 19, 2000 this sequence version replaced gi:8569912.

COMMENT
----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
Project Information
Center project name: H.NH0311D08
----- Summary Statistics -----
Sequencing vector: M13; 80%
Sequencing vector: plasmid; 20%
Chemistry: Dye-terminator ET; 80% of reads
Chemistry: Dye-terminator Big Dye; 20% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 181767 bases at least Q40
Consensus quality: 183800 bases at least Q30
Consensus quality: 185054 bases at least Q20
Insert size: 195000; agarose-fp
Insert size: 187098; sum-of-contigs
Quality coverage: 5.73 in Q20 bases; agarose-fp
Quality coverage: 5.24 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 9 contigs. The true order of the pieces
* is not known and their order in this sequence record is

```


GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: February 19, 2003, 14:53:22 : Search time 217.528 Seconds
(without alignments)
2084.671 Million cell updates/sec

Title: US-09-997-610-2_COPY_18_45
Perfect score: 162
Sequence: 1 AGPAPRRPEEVCPCGAPGCLPQYGEI 28

Scoring table:
BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 16154066 seqs, 809774376 residues
Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 200000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame_plus_p2n.model -DEV=x1h
-Q/cgn2_1/USPto_spool/US09997610/runat_10022003_160823_23843/app_query.fasta_1.1635
-DB=EST -QFMT=fastlap -SUPFI=1st -MINMATCH=0.1 -LOOPEXT=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=biolum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR SCORE=0.0 -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US09997610@cgn2_1.1871_etunat_10022003_160823_23843 -NCPU=6 -ICPU=3
-NO_XLPHY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV=TIMEOUT=120
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :
EST:
1: em_estba:*
2: em_estlum:*
3: em_estlin:*
4: em_estmu:*
5: em_estrov:*
6: em_estrpl:*
7: em_estro:*
8: em_hlc:*
9: gb_est1:*
10: gb_est2:*
11: gb_hlc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rtd:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	92	56.8	372	10	BB772103
2	88	54.3	447	9	AA934284
3	88	54.3	458	9	AA934212
4	88	54.3	531	9	AA934261
5	86	53.1	399	10	BE588347
6	86	53.1	445	9	AA857968
7	86	53.1	501	9	BE83103
8	86	53.1	524	12	BE846038
9	86	53.1	577	10	AV597164
10	86	53.1	662	10	AM257669
11	86	53.1	609	10	AM675980
12	85.5	52.8	626	10	AM675980
13	85.5	52.8	626	10	AM347951
14	85	52.5	391	14	H30955
15	85	52.5	490	9	AA585547
16	85	52.5	491	9	AA585556
17	85	52.5	563	10	AM348031
18	85	52.5	570	10	AM347952
19	85	52.5	571	10	AM225414
20	85	52.5	588	9	AA933455
21	85	52.5	598	10	AM675844
22	85	52.5	600	10	AM409438
23	85	52.5	637	10	AM682789
24	85	52.5	645	10	AM409436
25	85	52.5	645	10	AM409452
26	85	52.5	647	10	AM409421
27	85	52.5	650	10	AM409405
28	85	52.5	652	10	AM675873
29	85	52.5	653	10	AM497698
30	85	52.5	656	10	AM348029
31	85	52.5	657	10	AM225347
32	84	51.9	457	12	BE639395
33	84	51.9	565	9	AI515686
34	84	51.9	1101	17	CNS0024A
35	84	51.2	358	10	BE113822
36	83	51.2	388	9	AI444918
37	83	51.2	521	14	BQ266989
38	83	51.2	573	13	BE923552
39	83	51.2	599	10	BE306946
40	83	51.2	601	14	BQ266473
41	83	51.2	603	10	BE288346
42	83	51.2	666	14	BQ442816
43	83	51.2	674	14	BQ442305
44	83	51.2	726	14	BQ771348
45	83	51.2	738	12	BE864336

ALIGNMENTS

RESULT 1
BB772103
LOCUS
DEFINITION BB772103 RIKEN full-length enriched, B16 F10Y cells Mus musculus
ACCESSION BB772103
VERSION BB772103.1 GI:16213794
KEYWORDS
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
REFERENCE
1 (bases 1 to 372)
Akimura,T., Arahawa,T., Carinici,P., Furuno,M., Hanaagaki,T., Hayatsu,N., Hiramoto,K., Hirooka,T., Hirozane,T., Ishii

TITLE
JOURNAL
COMMENT

Unpublished (2001)
Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic
Sciences Center(GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Sushiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.go.jp,
URL: http://genome.gsc.riken.go.jp/
Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh
M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new
genes. Genome Res. 10 (10), 1617-1630 (2000)
Wagii, K., Fujiwaka, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E.,
Watanishi, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura
S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and
Hayashizaki, Y.
RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer. Genome Res.
10 (11), 1757-1771 (2000)
Konno, H., Fukunishi, Y., Shibata, K., Itoh, M., Carninci, P., Sugahara
Y. and Hayashizaki, Y.
Computer-based methods for the mouse full-length cDNA
encyclopedia: real-time sequence clustering for construction of a
nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
Please visit our web site (<http://genome.gsc.riken.go.jp>) for
further details.
e mouse tissues.
Location/Qualifiers
source
1. 372
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone_id="6370139P03"
/clone_lib="RIKEN full-length enriched, B16 F10Y cells"
/cell_type="B16 F10Y cells"
/note="Pooled tissues: (tissue_type=cerebellum,
dev_stage=16 days neonate, sex-mixed),
(tissue_type=cerebellum, dev_stage=0 day neonate,
sex-mixed), (tissue_type=hippocampus, dev_stage=adult,
sex-male), (tissue_type=whole body, dev_stage=9 days
embryo, sex-mixed), (tissue_type=lung, dev_stage=13 days
embryo, sex-mixed)"

BASE COUNT 69 a 138 c 90 g 75 t
ORIGIN

Alignment Scores:
Pred. No.: 14.9 Length: 372
Score: 92.00 Matches: 16
Percent Similarity: 81.82% Conservative: 2
Best Local Similarity: 72.73% Mismatches: 2
Query Match: 56.79% Indels: 1
DB: 10 Gaps: 1

US-09-997-610-2_COPY_18_45 (1-28) x BB772103 (1-372)

QY 3 PROPGALAHSPROAGPProGluGluValGlyProPro-----GlyAlaProGly 20
DB 65 CCCCAGGACATCTCGGCACCAAGGAACAGGCCCCAGACAAGGGGTGGCCAGGC 124

QY 21 LeuPro 22
DB 125 GTGCC 130

RESULT 2
AA934284
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

AA934284 447 bp mRNA linear EST 26-MAR-1999
SMOVL3CAN12F04 Onchocerca volvulus infective larva cDNA
(SAM94WL-OVL3) Onchocerca volvulus cDNA clone onch40 5' similar to
WP:R36A4.10 CE07185 CUTICULAR COLLAGEN :contains element PTR5
repetitive element ;, mRNA sequence.
AA934284
AA934284.1 GI:3091441
EST.
Onchocerca volvulus.
Onchocerca volvulus
Eukaryota; Metazoa; Nematoda; Chromadorea; Spirurida; Filarioidea;
Onchocercidae; Onchocerca.
1 (bases 1 to 447)
Williams, S.A., Lizotte-Waniewski, M., Laney, S., Wenhong, L., Hillier
L., Allen, M., Bowles, L., Geisel, S., Jost, S., Kuchta, T., Martin, J.,
Stephens, M., Theising, B., White, Y., Wylie, T., Chappell, J., Person, B.,
Gibbons, M., Harvey, N., Pape, D., Chamberlain, A., Morales, R.,
Schurk, R., Ritter, E., Kohn, S., Underwood, K. and Matra, M.
Molecular Parasitology OVL3
Unpublished (1998)
Contact: Steven A. Williams
Molecular Parasitology
Smith College Department of Biological Sciences
Department of Biological Sciences, Clark Science Center, Smith
College, Northampton, MA, 01063, USA
Tel: 4135853826
Fax: 4135853786
Email: genome@smith.edu
The library was constructed by Wenhong Lu. The library is available
from Dr. S.A. Williams, email genome@smith.edu When requesting this
clone from Dr. Williams, please reference the Williams lab clone id
- SMOVL3CAN12F04
Seq primer: -40ml3 fwd. RT from Amersham
High quality sequence stop: 432.
Location/Qualifiers
source
1. 447
/organism="Onchocerca volvulus"
/strain="Sierra Leone"
/db_xref="taxon:6282"
/clone="onch640"
/clone_lib="Onchocerca volvulus infective larva cDNA
(SAM94WL-OVL3)"
/lab_host="XLI-Blue MRP"
/note="Vector: lambda UniZap XR; Site 1: EcoR I; Site 2:
Xho I; Cutaneous filarial nematode parasite of humans
mRNA was prepared from third stage infective larvae of
Onchocerca volvulus isolated from mosquitoes 10 days after
infection and converted to double stranded cDNA using
reverse transcriptase and oligo(dT) followed by RNase H
and DNasep I. The library had 1.8 x 10⁵ independent
recombinants and average insert size was 900 base pairs.
The library was constructed by Wenhong Lu. The library is
available from Dr. S.A. Williams, email genome@smith.edu."

BASE COUNT 112 a 117 c 118 g 100 t
ORIGIN

Alignment Scores:
Pred. No.: 39.7 Length: 447
Score: 88.00 Matches: 16
Percent Similarity: 66.67% Conservative: 2
Best Local Similarity: 59.26% Mismatches: 9
Query Match: 54.32% Indels: 0
DB: 9 Gaps: 0

US-09-997-610-2_COPY_18_45 (1-28) x AA934284 (1-447)

QY 1 ALaGlyProGAlAHSPROAGPProGluGluValGlyProProGluAlaProGly 20
DB 304 GCGGACACACAGGTCTATCCCGGTAAACCTGTGTAAGCCAGGTAAACCTGTGAGCTCTGAC 363

Qy 21 LeuproglIntyThrglygu 27
|||||
Db 364 TTACCCGCAACCTGTAA 384

RESULT 3
LOCUS AA934212
DEFINITION SMOWL3CAN09C06 Onchocerca volvulus infective larva cDNA
(SAM94WL-OVL3) Onchocerca volvulus cDNA clone onch318 5' similar to
WP:C29F4.1 CE03038 COLLAGEN ; contains element PK5 repetitive
element ; mRNA sequence.

ACCESSION AA934212
VERSION AA934212.1 GI:3091369
KEYWORDS EST.
SOURCE Onchocerca volvulus.
ORGANISM Onchocerca volvulus.
Eukaryota; Metazoa; Nematoda; Chromadorea; Spirurida; Filarioidea;
Onchocercidae; Onchocerca.

REFERENCE 1 (bases 1 to 458)
AUTHORS Williams,S.A., Lizotte-Waniewski,M., Laney,S., Wenhong,L., Hillier,
L., Allen,M., Bowles,L., Gelsel,S., Jost,S., Kucaba,T., Martin,J.,
Stephens,M., Theising,B., White,Y., Wylie,T., Chappell,J., Person,B.,
Gibbons,M., Harvey,N., Pape,D., Chamberlain,A., Morales,R.,
Schurk,R., Ritter,E., Kohn,S., Underwood,K. and Marra,M.
Molecular Parasitology OVL3
Unpublished (1998)
Contact: Steven A. Williams
Molecular Parasitology
Smith College Department of Biological Sciences
Department of Biological sciences, Clark Science Center, Smith
College, Northampton, MA, 01063, USA
Tel: 4135853826
Fax: 4135853786
Email: genome@smith.edu
The library was constructed by Wenhong Lu. The library is available
from Dr. S.A. Williams, email genome@smith.edu When requesting this
clone from Dr. Williams, please reference the Williams lab clone id
- SMOWL3CAN09C06
Seq primer: -40m13 fwd. ET from Amersham
High quality sequence stop: 456.
Location/Qualifiers
1. 458
/organism="Onchocerca volvulus"
/strain="Sierra Leone"
/db_xref="taxon:6282"
/clone="onch318"
/clone_11b="Onchocerca volvulus infective larva cDNA
(SAM94WL-OVL3)"
/lab_host="XLI-Blue MRF"
/note="Vector: lambda Unizap XR; Site_1: Ecor I; Site_2:
Xho I; Cutaneous filarial nematode parasite of humans.
mRNA was prepared from third stage infective larvae of
Onchocerca volvulus isolated from mosquitoes 10 days after
infection and converted to double stranded cDNA using
reverse transcriptase and oligo(dT) followed by RNase H
and DNasep I. The library had 1.8 x 10⁵ independent
recombinants and average insert size was 900 base pairs.
The library was constructed by Wenhong Lu. The library is
available from Dr. S.A. Williams, email genome@smith.edu."

BASE COUNT 113 a 122 c 121 g 102 t

ORIGIN

Alignment Scores:
Pred. No.: 40.5 Length: 458
Score: 88.00 Matches: 16
Percent Similarity: 66.67% Conservative: 2
Best Local Similarity: 59.26% Mismatches: 9
Query Match: 54.32% Indels: 0
DB: 9 Gaps: 0

US-09-997-610-2_COPY_18_45 (1-28) x AA934212 (1-458)

Qy 1 AtaglyProProLahlsProaArgProProGLUValGlyProProGLyAlaProGLy 20

Db 311 GCGGACACACAGTCAACCCGGTAACCTGTGAACGAGTAACCTGAGCTCTGGC 370
|||||

Qy 21 LeuproglIntyThrglygu 27
|||||
Db 371 TTACCCGCAACCTGTAA 391

RESULT 4
LOCUS AA934261
DEFINITION SMOWL3CAN12B08 Onchocerca volvulus infective larva cDNA
(SAM94WL-OVL3) Onchocerca volvulus cDNA clone onch596 5' similar to
WP:F364L.10 CE07185 CUTICULAR COLLAGEN ; contains element MS1
repetitive element ; mRNA sequence.

ACCESSION AA934261
VERSION AA934261.1 GI:3091418
KEYWORDS EST.
SOURCE Onchocerca volvulus.
ORGANISM Onchocerca volvulus.
Eukaryota; Metazoa; Nematoda; Chromadorea; Spirurida; Filarioidea;
Onchocercidae; Onchocerca.

REFERENCE 1 (bases 1 to 531)
AUTHORS Williams,S.A., Lizotte-Waniewski,M., Laney,S., Wenhong,L., Hillier,
L., Allen,M., Bowles,L., Gelsel,S., Jost,S., Kucaba,T., Martin,J.,
Stephens,M., Theising,B., White,Y., Wylie,T., Chappell,J., Person,B.,
Gibbons,M., Harvey,N., Pape,D., Chamberlain,A., Morales,R.,
Schurk,R., Ritter,E., Kohn,S., Underwood,K. and Marra,M.
Molecular Parasitology OVL3
Unpublished (1998)
Contact: Steven A. Williams
Molecular Parasitology
Smith College Department of Biological Sciences
Department of Biological sciences, Clark Science Center, Smith
College, Northampton, MA, 01063, USA
Tel: 4135853826
Fax: 4135853786
Email: genome@smith.edu
The library was constructed by Wenhong Lu. The library is available
from Dr. S.A. Williams, email genome@smith.edu When requesting this
clone from Dr. Williams, please reference the Williams lab clone id
- SMOWL3CAN12B08
Seq primer: -40m13 fwd. ET from Amersham
High quality sequence stop: 485.
Location/Qualifiers
1. 531
/organism="Onchocerca volvulus"
/strain="Sierra Leone"
/db_xref="taxon:6282"
/clone="onch596"
/clone_11b="Onchocerca volvulus infective larva cDNA
(SAM94WL-OVL3)"
/lab_host="XLI-Blue MRF"
/note="Vector: lambda Unizap XR; Site_1: Ecor I; Site_2:
Xho I; Cutaneous filarial nematode parasite of humans.
mRNA was prepared from third stage infective larvae of
Onchocerca volvulus isolated from mosquitoes 10 days after
infection and converted to double stranded cDNA using
reverse transcriptase and oligo(dT) followed by RNase H
and DNasep I. The library had 1.8 x 10⁵ independent
recombinants and average insert size was 900 base pairs.
The library was constructed by Wenhong Lu. The library is
available from Dr. S.A. Williams, email genome@smith.edu."

BASE COUNT 142 a 130 c 139 g 115 t

ORIGIN

Alignment Scores:
Pred. No.: 46 Length: 531
Score: 88.00 Matches: 16
Percent Similarity: 66.67% Conservative: 2
Best Local Similarity: 59.26% Mismatches: 9
Query Match: 54.32% Indels: 0
DB: 9 Gaps: 0

KEYWORDS	EST.
SOURCE	cow.
ORGANISM	Bos taurus
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidea; Bovidae; Bovinae; Bos.
AUTHORS	1 (bases 1 to 499) Sonstegard,T.S., Capuco,A.V., Van Tassel,C.P., Ashwell,M.S. and Wells,K.D.
TITLE	Mapping of Expressed Sequence Tags from a normalized bovine mammary gland cDNA library
JOURNAL	Unpublished (2000)
COMMENT	Contact: Sonstegard TS USA, ARS, Beltsville Agricultural Research Center Bldg. 200, Rm 2a, Beltsville, MD 20705, USA Tel: 301 504 8416 Fax: 301 504 8414 Email: tads@nrl.barc.usda.gov Single pass sequencing. Bases called and alt. trimmed with phred v0.980904.e. Vector identified by cross_match with the -mismatch 18 and -mismatch 12 options. PCR primers FORWARD: AGGAACAGCTATGACCAT BACKWARD: GTTGTCCAGTCACGACG Plate: 13 row: B column: 18 Seq primer: ATTAGCGACACTATGAC.
FEATURES	Seq primer: ATTAGCGACACTATGAC. Location/Qualifiers 1..499 /organism="Bos taurus" /db_xref="taxon:9913" /clone_lib="BARC 5BOY" /tissue_type="pooled" /lab_host="DH10B" /note="Vector: PCMV SPORT6, site_1: NotI, site_2: SalI; Library made from pooled mRNA isolated from mammary tissues at eight physiological, developmental, and disease states."
BASE COUNT	105 a 160 c 129 g 105 t
ORIGIN	
Alignment Scores:	
Pred. NO.:	65.7 Length: 499
Score:	86.00 Matches: 16
Percent Similarity:	61.54% Conservative: 0
Best Local Similarity:	61.54% Mismatches: 10
Query Match:	53.09% Indels: 0
DB:	10 Gaps: 0
US-09-997-610-2_COPY_18_45 (1-28) x BE483103 (1-499)	
QY 1	AAGAGTProfoALnHsProArgProfoGluGluValGlyProfoGlyALaProGly 20
Db 386	GCTGCTCTCTCGGCGCCACCGACCCCTGGACATCTGGCATCTCTGTGCCCTGCGC 445
Y 21	LeuProGlnTyrThnGly 26
Db 446	GCTCCAGATACCAAGT 463
RESULT 8	
AA847768/c	
LOCUS	AA847768 501 bp mRNA EST 31-MAR-1998
DEFINITION	oe9107.51 NCI-CGAP.C012 Homo sapiens cDNA clone IMAGE:1419781
ACCESSION	AA847768
VERSION	AA847768.1 GI:2934286
KEYWORDS	EST.
SOURCE	human.
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
AUTHORS	NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.

JOURNAL	National Cancer Institute, Cancer Genome Anatomy Project (CGAP),	
COMMENT	Tumor Gene Index Unpublished (1997) Contact: Robert Strausberg, Ph.D. Email: cgapbs-r@mail.nih.gov Tissue Procurement: L. Jeffrey Medeiros, M.D., Michael R. Emmert-Buck, M.D., Ph.D. CDNA Library Preparation: Stratagene, Inc. CDNA Library Arrayed by: Greg Lennon, Ph.D. DNA Sequencing by: Washington University Genome Sequencing Center Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/ILN/ILN at: www-bio.llnl.gov/bbrp/image/image.html Insert Length: 1068 Std Error: 0.00 Seq primer: -40m13 fwd. Et from Amersham High quality sequence stop: 339. Location/Qualifiers 1..501 /organism="Homo sapiens" /db_xref="taxon:9606" /clone_image:1419781" /clone_id="NCI_CGAP_C012" /sex="mixed" /tissue_type="colon tumor" /lab_host="SOLR (kanamycin resistant)" /note="Organ: colon; Vector: Bluescript SK-; Site_1: EcoRI / Site_2: XhoI; Cloned unidirectionally. Primer: Oligo dt. Pooled colon tumors. 5' adaptor sequence: 5' GAATTGGCAGCAG 3' 3' adaptor sequence: 5' CTCAGTTTTTTTTTTTTTTTTTTT 3' Average insert size: 1.2 kb." BASE COUNT 91 a 126 c 186 g 98 t ORIGIN	
FEATURES	source	
ALIGNMENT	Alignment Scores: Pred. No.: 65.9 Length: 501 Score: 86.00 Matches: 16 Percent Similarity: 73.91% Conservative: 1 Best Local Similarity: 69.57% Mismatch: 4 Query Match: 53.09% Indels: 2 DB: 9 Gaps: 1	
US-09-997-610-2_COPY_18_45 (1-28) x AA847768 (1-501)		
Oy 2	GIYPROROLANLSPATGTRGROGUGUVALGIYPRORP-----GIYALAPRO 19	
Db 333	GGACCCSCAGGCCSCAGGCCSCCAAGAGCAGGCCGCCSCAGACAGGATGGCCG 274	
Oy 20	GIYLeuPro 22	
Db 273	GGCCGCCCC 265	
RESULT 9		
BE846038		
LOCUS	BE846038 524 bp mRNA linear EST 25-SEP-2000	
DEFINITION	233389 BAKC 5BOV Bos taurus cDNA 5', mRNA sequence.	
ACCESSION	BE846038	
VERSION	BE846038.1 GI:10282862	
KEYWORDS	EST.	
SOURCE	CCV.	
ORGANISM	Bos taurus	
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidea; Bovidae; Bovinae; Bos. 1 (bases 1 to 524) Sonstegard,T.S., Capuco,A.V., Van Tassel,C.P., Ashwell,M.S. and Wells,K.D.	
TITLE	Mapping of Expressed Sequence Tags from a normalized bovine mammary gland cDNA library Unpublished (2000) Contact: Sonstegard TS USA, ARS, Beltsville Agricultural Research Center Bd19, 200 Km 2A, Beltsville, MD 20705, USA Tel: 301 504 8416	
JOURNAL		
COMMENT		

Fax: 301 504 8414
Email: tadsaentri.barc.usda.gov
Single pass sequencing. Bases called and alt-trimmed with phred v0.980904.e. Vector identified by cross_match with the -minscore 18 and -mismatch 12 options.
PCR primers
FORWARD: AGGACACGCTATGACCAT
BACKWARD: GTTTCCTCCAGTCACGACG
Plate: 114 row: F column: 10
Seq primer: ATTTAGTGCACCTATAG.
Location/Qualifiers
1. 524
/organism="Bos taurus"
/db_xref="taxon:9913"
/clone_1id="BARC 5BOV"
/tissue_type="pooled"
/lab_host="DH10B"
/note="Vector: PCMV SPORT6; site_1: NotI; site_2: SalI;
library made from pooled mRNA isolated from mammary
tissues at eight physiological, developmental, and disease
states."
BASE COUNT 119 a 139 c 167 g 99 t
ORIGIN

Alignment Scores:
Pred. No.: 68.5 Length: 524
Score: 86.00 Matches: 16
Percent Similarity: 61.54% Conservative: 0
Best Local Similarity: 61.54% Mismatches: 10
Query Match: 53.09% Indels: 0
DB: 12 Gaps: 0

US-09-997-610-2_COPY_18_45 (1-28) x BE846038 (1-524)
QY 1 AAlaGyPProGAlaHisPProArGProProGluGluValGlyProProGlyAlaProGly 20
|||||
Db 113 GGTGGTCTCTCTGGCCGCCGACCCCTGGCACATCTGGCCATCTGGTCCCTGGC 172
QY 21 LeuProGlnTyThrGly 26
||| ||| |||
Db 173 GCTCCAGGATACCAAGT 190

RESULT 10
AV597164 577 bp mRNA linear EST 27-NOV-2001
LOCUS AV597164 Bos taurus cartilage fetus Bos taurus cDNA clone
DEFINITION E1CA030C06 5', mRNA sequence.
ACCESSION AV597164
VERSION AV597164.1 GI:9714506
KEYWORDS EST.
SOURCE
ORGANISM Bos taurus
COV.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovidae; Bovinae; Bos.
REFERENCE 1 (bases 1 to 577)
AUTHORS Takasuga,A., Hirotsume,S., Itoh,R., Itohzono,A., Suzuki,H., Aso,H.
and Sugimoto,Y.
TITLE Establishment of a high throughput EST sequencing system using
poly(A) tail-removed cDNA libraries and determination of 36,000
bovine ESTs
Nucleic Acids Res. 29 (22), E108 (2001)
21570554
Contact: Yoshikazu Sugimoto
Animal Genetics Division
Shirakawa Institute of Animal Genetics
OdaKura, Nishigo, Nishi-Shirakawa, Fukushima 961-8061, Japan
Tel: 81-248-25-5641
Fax: 81-248-25-5725
Email: kazusugieccococoon.ne.jp
Single pass sequencing.
This clone was obtained from a polyA-deleted cDNA library.
Location/Qualifiers

FEATURES

FEATURES

Location/Qualifiers

source
1. 577
/organism="Bos taurus"
/db_xref="taxon:9913"
/clone="E1CA030C06"
/clone_1id="Bos taurus cartilage fetus"
/tissue_type="cartilage"
/dev_stage="fetus"
/lab_host="DH10B"
/note="Vector: pZL1; site_1: SalI; site_2: NotI; Poly A
was deleted from a NotI site"
BASE COUNT 123 a 170 c 171 g 113 t
ORIGIN

Alignment Scores:
Pred. No.: 74.5 Length: 577
Score: 86.00 Matches: 16
Percent Similarity: 61.54% Conservative: 0
Best Local Similarity: 53.09% Mismatches: 10
Query Match: 53.09% Indels: 0
DB: 10 Gaps: 0

US-09-997-610-2_COPY_18_45 (1-28) x AV597164 (1-577)
QY 1 AAlaGyPProGAlaHisPProArGProProGluGluValGlyProProGlyAlaProGly 20
|||||
Db 245 GGTGGTCTCTCTGGCCGCCGACCCCTGGCACATCTGGCCATCTGGTCCCTGGC 304
QY 21 LeuProGlnTyThrGly 26
||| ||| |||
Db 305 GCTCCAGGATACCAAGT 322

RESULT 11
AW257669 662 bp mRNA linear EST 20-DEC-1999
LOCUS AW257669
DEFINITION SWYD25CAU08D04K Brugia malayi young adult day 25 cDNA
(SAM99MLW-BMYD25) Brugia malayi cDNA clone SWYD25CAU08D04 5', mRNA
sequence.
ACCESSION AW257669
VERSION AW257669.1 GI:6605842
KEYWORDS EST.
SOURCE Brugia malayi.
ORGANISM Brugia malayi.
Eukaryota; Metazoa; Nematoda; Chromadorea; Spirurida; Filarioidae;
Onchocercidae; Brugia.
REFERENCE 1 (bases 1 to 662)
AUTHORS Williams,S.A.
TITLE Genes expressed in young adult day 25 of Brugia malayi
JOURNAL Unpublished (1999)
COMMENT Contact: Steven A. Williams
Molecular Parasitology
Smith College Department of Biological Sciences
Department of Biological Sciences, Clark Science Center, Smith
College, Northampton, MA, 01063, USA
Tel: 4135853826
Fax: 4135853786
Email: genomes@smith.edu
Seq primer: pBluescript SK.
Location/Qualifiers
1. 662
/organism="Brugia malayi"
/db_xref="taxon:6279"
/clone="SWYD25CAU08D04"
/clone_1id="Brugia malayi young adult day 25 cDNA
(SAM99MLW-BMYD25)"
/dev_stage="young adult, twenty five days after infection"
/lab_host="XLI-Blue MRF"
/note="Vector: Lambda Uni-ZAP XR; Site_1: Eco RI; Site_2:
Xho I; Lymphatic filarial nematode parasite of humans.
mRNA was prepared from young adult worms isolated from
the peritoneal cavity of jirds on day 25 after infection
and converted to double-stranded cDNA using reverse
transcriptase and oligo(dT) followed by RNase H and DNA
pol I. The library has 6.2 x 10⁵ independent recombinants

FEATURES

source

and the average insert size is approx.1101bp. The library was constructed by Michelle Lizotte-Waniewski. The library is available from Dr. S.A. Williams, email: genome@neal.smith.edu."

BASE COUNT 195 a 176 c 165 g 120 t 6 others

ORIGIN

Alignment Scores:

Pred. No.:	83.8	Length:	662
Score:	86.00	Matches:	18
Percent Similarity:	54.55%	Conservative:	0
Best Local Similarity:	54.35%	Mismatches:	9
Query Match:	53.09%	Indels:	6
DB:	10	Gaps:	1

US-09-997-610-2_COPY_18_45 (1-28) x AW257669 (1-662)

OY 1 AAlAGlyProProAlahisProArg-----ProProGluGluValGly 14

DB 458 GCCGGACACCTGGCAGACCAGCAATAGTGCATGTCTGCACACGAGACCGTGGGA 517

OY 15 PROProGlyAlaProGlyLeuProGIntYrThGlyGlu 27

DB 518 CCACGAGAGACACGAGTCCAGCCAGGAGCTGACGCGCAA 556

RESULT 12

AW675980 609 bp mRNA linear EST 27-APR-2000

LOCUS SWD25CAU013F09SK Brugia malayi young adult day 25 CDNA

DEFINITION (SAM99MLW-BmyD25) Brugia malayi cDNA clone SWD25CAU013F09 5', mRNA

sequence.

ACCESSION AW675980

VERSION AW675980

KEYWORDS EST.

SOURCE Brugia malayi.

ORGANISM Brugia malayi.

REFERENCE 1 (bases 1 to 609)

AUTHORS Williams,S.A.

TITLE Genes expressed in young adult day 25 of Brugia malayi

JOURNAL Unpublished (1999)

COMMENT Molecular Parasitology

Smith College Department of Biological Sciences

Department of Biological Sciences, Clark Science Center, Smith

College, Northampton, MA, 01063, USA

Tel: 4135853826

Fax: 4135853786

Email: genome@smith.edu

Seq primer: plusescript SK.

Location/Qualifiers

1. 609

/organism="Brugia malayi"

/db_xref="taxon:6279"

/clone="SWD25CAU013F09"

/clone_lib="Brugia malayi young adult day 25 CDNA

Unpublished (1999)

COMMENT Molecular Parasitology

Smith College Department of Biological Sciences

Department of Biological Sciences, Clark Science Center, Smith

College, Northampton, MA, 01063, USA

Tel: 4135853826

Fax: 4135853786

Email: genome@smith.edu

Seq primer: plusescript SK.

Location/Qualifiers

1. 626

/organism="Brugia malayi"

/db_xref="taxon:6279"

/clone="SWD25CAU09B11"

/clone_lib="Brugia malayi young adult day 25 CDNA

(SAM99MLW-BmyD25)"

/dev_stage="young adult, twenty five days after infection"

/lab_host="XLI-Blue MRF"

/note="Vector: Lambda Uni-ZAP XR; Site_1: Eco RI; Site_2:

Xho I; Lymphatic filarial nematode parasite of humans.

mRNA was prepared from young adult worms isolated from

the peritoneal cavity of jirds on day 25 after infection

and converted to double-stranded cDNA using reverse

transcriptase and oligo(dT) followed by RNase H and DNA

pol I. The library has 6.2 x 105 independent recombinants

and the average insert size is approx.1101bp. The library

was constructed by Michelle Lizotte-Waniewski. The

library is available from Dr. S.A. Williams, email:

genome@neal.smith.edu."

BASE COUNT 189 a 164 c 159 g 114 t

ORIGIN

Alignment Scores:

Pred. No.:

Score:

Percent Similarity:

Best Local Similarity:

Query Match:

DB:

Alignment Scores:

Pred. No.:

Score:

Percent Similarity:

Best Local Similarity:

Query Match:

DB:

US-09-997-610-2_COPY_18_45 (1-28) x AW675980 (1-609)

OY 1 AAlAGlyProProAlahisPro-----ArgProProGluGluValGlyPro 15

DB 439 GCCGGACACCTGGCAGACCAGCAATAGTGCATGTCTGCACACGAGACCGTGGACA 498

OY 16 PROGlyAlaProGlyLeuProGIntYrThGlyGlu 27

DB 499 CCAGGAGACACGAGTCCAGCCAGGAGCTGACGCGCAA 534

RESULT 13

AW347951 626 bp mRNA linear EST 01-FEB-2000

LOCUS SWD25CAU09B11SK Brugia malayi young adult day 25 CDNA

DEFINITION (SAM99MLW-BmyD25) Brugia malayi cDNA clone SWD25CAU09B11 5', mRNA

sequence.

ACCESSION AW347951

VERSION AW347951

KEYWORDS EST.

SOURCE Brugia malayi.

ORGANISM Brugia malayi.

REFERENCE 1 (bases 1 to 626)

AUTHORS Williams,S.A.

TITLE Genes expressed in young adult day 25 of Brugia malayi

JOURNAL Unpublished (1999)

COMMENT Molecular Parasitology

Smith College Department of Biological Sciences

Department of Biological Sciences, Clark Science Center, Smith

College, Northampton, MA, 01063, USA

Tel: 4135853826

Fax: 4135853786

Email: genome@smith.edu

Seq primer: plusescript SK.

Location/Qualifiers

1. 626

/organism="Brugia malayi"

/db_xref="taxon:6279"

/clone="SWD25CAU09B11"

/clone_lib="Brugia malayi young adult day 25 CDNA

(SAM99MLW-BmyD25)"

/dev_stage="young adult, twenty five days after infection"

/lab_host="XLI-Blue MRF"

/note="Vector: Lambda Uni-ZAP XR; Site_1: Eco RI; Site_2:

Xho I; Lymphatic filarial nematode parasite of humans.

mRNA was prepared from young adult worms isolated from

the peritoneal cavity of jirds on day 25 after infection

and converted to double-stranded cDNA using reverse

transcriptase and oligo(dT) followed by RNase H and DNA

pol I. The library has 6.2 x 105 independent recombinants

and the average insert size is approx.1101bp. The library

was constructed by Michelle Lizotte-Waniewski. The

library is available from Dr. S.A. Williams, email:

genome@neal.smith.edu."

BASE COUNT 189 a 164 c 159 g 114 t

ORIGIN

Alignment Scores:

Pred. No.:

Score:

Percent Similarity:

Best Local Similarity:

Query Match:

DB:

Alignment Scores:

Pred. No.:

Score:

Percent Similarity:

Best Local Similarity:

Query Match:

DB:

Alignment Scores:

Pred. No.:

Score:

Percent Similarity:

Best Local Similarity:

Query Match:

DB:

Qy 15 ProProGlyAlaProGlyLeuProGlnTyrThrGlyGlu 27
|||||
Db 66 CCACCAGAGACACACAGTCAGCCAGGAGCTGAGAGGGGAA 104
|||||

Search completed: February 19, 2003, 19:36:01
Job time : 219.528 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: February 19, 2003, 14:57:18 ; Search time 390.136 Seconds
(without alignments)
2088.706 Million cell updates/sec

Title: US-09-997-610-2_COPY_18_45
Perfect score: 162
Sequence: 1 AGPPHPPREVEVPPGAPGLPYTGEI 28

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2054640 seqs, 14551402878 residues
Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL=frame+_p2n.model -DEV_xlh
-O/cgn2.1/USPTO/spool/US09997610/runal_10022003_160823_23819/app_query.fasta.1.1635
-DB=genEmbl -QFMT=fastap -SUFFIX=rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blsum62 -TRANS=human4.0.cdi -LIST=45
-DOCALLIGN=200 -THR.SCORE=pct -THR.MAX=100 -THR.MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=fto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US09997610_eCGN_1_1_3955_erunal_10022003_160823_23819 -NCPU=6 -ICPU=3
-NO_XLPUT -NO_MMAP -LARGEQUERY -NEG.SCORES=0 -WAIT -LONGLOG -DEV.TIMEOUT=120
-WARN.TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : GenEmbl:*
1: gb_da:*
2: gb_hcg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*
15: em_da:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*

29: em_vl:*
30: em_hcg_hum:*
31: em_hcg_inv:*
32: em_hcg_other:*
33: em_hcg_mus:*
34: em_hcg_pin:*
35: em_hcg_frod:*
36: em_hcg_mam:*
37: em_hcg_vrt:*
38: em_sy:*
39: em_hcgo_hum:*
40: em_hcgo_mus:*
41: em_hcgo_other:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	119	73.5	145880	9 HS302D9	Z82198 Human DNA s
2	86	53.1	4428	6 AX146422	AX146422 Sequence
3	86	53.1	4428	6 AX146424	AX146424 Sequence
4	84.5	52.2	68661	3 AC024805	AC024805 Caenorhab
5	84.5	52.2	278007	2 AC006799	AC006799 Caenorhab
6	84	51.9	4428	6 AX146430	AX146430 Sequence
7	84	51.9	66669	3 AF017777	AF017777 Drosophila
8	84	51.9	87089	2 AC013980	AC013980 Drosophila
9	84	51.9	171548	3 AC092494	AC092494 Drosophila
10	84	51.9	173508	3 AC011251	AC011251 Drosophila
11	84	51.9	256417	2 AC079422	AC079422 Mus muscu
12	84	51.9	292911	3 AE003568	AE003568 Drosophila
13	83	51.2	186	6 AR014097	AR014097 Sequence
14	83	51.2	186	6 AR117066	AR117066 Sequence
15	83	51.2	186	6 I06482	I06482 Sequence 19
16	83	51.2	186	6 I20524	I20524 Sequence 62
17	83	51.2	287	3 O05420010	AJ420010 Osterlag1
18	83	51.2	3612	10 NMICOLA4	X05777 Mouse mRNA
19	83	51.2	6512	6 AX306181	AX306181 Sequence
20	83	51.2	6512	10 NMICOLA4A	J04694 Mus musculus
21	82.5	50.9	437	10 NMMA379COL	X91012 M. musculus
22	82.5	50.9	1009	9 BC004412	BC004412 Homo sapi
23	82.5	50.9	1521	10 AF237721	AF237721 Mus muscu
24	82.5	50.9	1605	10 BC030945	BC030945 Mus muscu
25	82.5	50.9	2091	9 BC007574	BC007574 Homo sapi
26	82.5	50.9	2388	10 AF349718	AF349718 Mus muscu
27	82.5	50.9	2437	9 BC011705	BC011705 Homo sapi
28	82.5	50.9	2480	9 HUMCOL31X	L41162 Homo sapien
29	82.5	50.9	2543	6 AR014074	AR014074 Sequence
30	82.5	50.9	2543	6 AR111836	AR111836 Sequence
31	82.5	50.9	6246	9 HSCOL9A3S2	AF026802 Homo sapi
32	82.5	50.9	69252	2 AC101519	AC101519 Mus muscu
33	82.5	50.9	160241	9 HS88517	AI035669 Human DNA
34	82.5	50.9	226554	10 AL669926	AL669926 Mouse DNA
35	82	50.6	426	9 AF230926	AF230926 Macaca mu
36	82	50.6	994	6 AX067354	AX067354 Sequence
37	82	50.6	1485	10 M0SC3A1A	M18933 Mouse alpha
38	82	50.6	1881	6 A60690	A60690 Sequence 20
39	82	50.6	1881	6 AR123971	AR123971 Sequence
40	82	50.6	3234	9 HSC3A1R	X15332 Human COL3A
41	82	50.6	3902	9 BC028178	BC028178 Homo sapi
42	82	50.6	5460	6 AX302553	AX302553 Sequence
43	82	50.6	5460	6 AX393295	AX393295 Sequence
44	82	50.6	5460	6 E10600	E10600 cDNA encod1
45	82	50.6	5460	9 HSCOL3A1	X14420 Human mRNA

ALIGNMENTS

RESULT 1

H30209
LOCUS H30209 145880 bp DNA linear PRI 12-DEC-1999
DEFINITION Human DNA sequence from clone RP1-302D9 on chromosome 22 Contains
GSSS, complete sequence.
ACCESSION 282198
VERSION 282198.2 GI:6572207
KEYWORDS HMG.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 145880)
AUTHORS Bridgeman, A.
TITLE Direct Submission
JOURNAL Submitted (08-DEC-1999) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
On Dec 13, 1999 this sequence version replaced gi:3164067.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence has been finished according to sequence map criteria
as follows. An attempt is made to resolve all sequencing problems,
such as compressions and repeats, but not necessarily within known
annotated human repeat sequence elements (e.g. Alu). Where the
sequence is ambiguous, there is an annotation using the 'unsure'
feature key.
The following abbreviations are used to associate primary accession
numbers given in the feature table with their source databases:
Em, EMBL; SW, SWISSPROT; Tr, TREMBL; Wp, WORMEP; Information
at http://www.sanger.ac.uk/Projects/C_elegans/wormep This sequence
was generated from part of bacterial clone contigs of human
chromosome 22, constructed by the Sanger Centre Chromosome 22
Mapping Group. Further information can be found at
<http://www.sanger.ac.uk/HGP/Chr22>
RP1-302D9 is from the library RPC1-1 constructed at the Roswell
Park Cancer Institute by the group of Pieter de Jong. For further
details see <http://bacpac.med.buffalo.edu/>
VECTOR: pCYPAC2
This sequence is the entire insert of clone RP1-302D9 The true left
end of clone CTA-282E2 is at 69682 in this sequence. The true right
end of clone CTA-41562 is at 55167 in this sequence.
FEATURES
source
1..145880
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="22"
/clone="RP1-302D9"
/clone_1lb="RPC1-1"
188..245
/note="MER3 repeat: matches 144. .209 of consensus"
246..571
/note="AluX repeat: matches 1. .312 of consensus"
572..759
/note="MER3 repeat: matches 1. .144 of consensus"
763..933
/note="MER5A repeat: matches 26. .187 of consensus"
1033..1336
/note="AluSP repeat: matches 1. .299 of consensus"
1450..1583
/note="MIR repeat: matches 24. .160 of consensus"
1687..1752
/note="L2 repeat: matches 2593. .2661 of consensus"
2350..2660
/note="AluSC repeat: matches 3. .309 of consensus"
2684..2981
/note="AluSq repeat: matches 2. .300 of consensus"
3323..3343
/note="MUTIE repeat: matches 116. .136 of consensus"

repeat_region 3344..3652
/note="AluY repeat: matches 1. .309 of consensus"
repeat_region 3653..3928
/note="MUTIE repeat: matches 136. .359 of consensus"
repeat_region 3929..4278
/note="MERB repeat: matches 3. .364 of consensus"
repeat_region 4279..4485
/note="MUTIE repeat: matches 359. .568 of consensus"
repeat_region 5073..5176
/note="52 copies 2 mer ct 78 conserved"
5181..5491
/note="AluDb repeat: matches 1. .311 of consensus"
6369..6483
/note="L2 repeat: matches 2579. .2705 of consensus"
6647..6685
/note="WAD1 repeat: matches 1. .23 of consensus"
6686..6987
/note="AluSX repeat: matches 1. .302 of consensus"
6988..7036
/note="WAD1 repeat: matches 23. .77 of consensus"
7482..7754
/note="AluDb repeat: matches 9. .290 of consensus"
7775..8060
/note="AluJo repeat: matches 1. .295 of consensus"
8414..8551
/note="L2 repeat: matches 2553. .2706 of consensus"
8914..9030
/note="MIR repeat: matches 147. .262 of consensus"
9110..9280
/note="MIR repeat: matches 91. .262 of consensus"
9283..9412
/note="MIR repeat: matches 15. .144 of consensus"
9521..9679
/note="FAM repeat: matches 3. .161 of consensus"
9820..10225
/note="MSTRB repeat: matches 2. .425 of consensus"
complement(10179..10678)
/note="match: GSS: Em:856592"
complement(10204..10728)
/note="match: GSS: Em:AQ071486"
complement(10249..10706)
/note="match: GSS: Em:AQ225455"
10312..10383
/note="MIR repeat: matches 79. .150 of consensus"
10718..11310
/note="match: GSS: Em:B14024"
10784..11201
/note="match: GSS: Em:B43656"
11838..11946
/note="MIR repeat: matches 20. .137 of consensus"
12174..12445
/note="L2 repeat: matches 1988. .2275 of consensus"
12444..12642
/note="MIR repeat: matches 63. .241 of consensus"
13017..13369
/note="match: STRS: Em:G49301"
13331..13397
/note="MIR repeat: matches 174. .244 of consensus"
13398..13698
/note="AluSP repeat: matches 1. .302 of consensus"
13699..13810
/note="MIR repeat: matches 76. .174 of consensus"
13806..13919
/note="MIR repeat: matches 77. .189 of consensus"
13945..14060
/note="MIR repeat: matches 24. .142 of consensus"
14061..14367
/note="AluY repeat: matches 1. .301 of consensus"
14368..14452
/note="MIR repeat: matches 141. .225 of consensus"
14589..14679
/note="MIR repeat: matches 173. .262 of consensus"
14597..15201
misc_feature

[illegible]

DEFINITION	Sequence 3 from Patent WO0134647.
ACCESSION	AX146422
VERSION	AX146422.1
KEYWORDS	GI:14284845
SOURCE	.
ORGANISM	cow.
	Bos taurus
	Eumariota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidea; Bovidae; Bovinae; Bos.
REFERENCE	1 (bases 1 to 4428)
AUTHORS	Bell,M.P., Neff,T.B., Polarek,J.W. and Seeley,T.W.
TITLE	Animal collagens and gelatins
JOURNAL	Patent: WO 0134647-A 3 17-MAY-2001; FIBROGEN, INC. (US)
FEATURES	Location/Qualifiers
source	1..4428 /organism="Bos taurus" /db_xref="taxon:9913"
BASE COUNT	973 a 1185 c 1400 g 870 t
ORIGIN	
Alignment Scores:	
Pred. No.:	89.9
Score:	86.00
Percent Similarity:	61.54%
Best Local Similarity:	61.54%
Query Match:	53.09%
DB:	6
	Gaps:
	0
US-09-997-610-2_COPY_18_45 (1-28) x AX146422 (1-4428)	
OY	1
DB	538
OY	21
DB	598
RESULT 3	
LOCUS	AX146424
DEFINITION	Sequence 5 from Patent WO0134647.
ACCESSION	AX146424
VERSION	AX146424.1
KEYWORDS	GI:14284846
SOURCE	.
ORGANISM	cow.
	Bos taurus
	Eumariota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidea; Bovidae; Bovinae; Bos.
REFERENCE	1 (bases 1 to 4428)
AUTHORS	Bell,M.P., Neff,T.B., Polarek,J.W. and Seeley,T.W.
TITLE	Animal collagens and gelatins
JOURNAL	Patent: WO 0134647-A 5 17-MAY-2001; FIBROGEN, INC. (US)
FEATURES	Location/Qualifiers
source	1..4428 /organism="Bos taurus" /db_xref="taxon:9913"
BASE COUNT	974 a 1185 c 1399 g 870 t
ORIGIN	
Alignment Scores:	
Pred. No.:	89.9
Score:	86.00
Percent Similarity:	61.54%
Best Local Similarity:	61.54%
Query Match:	53.09%
DB:	6
	Gaps:
	0
US-09-997-610-2_COPY_18_45 (1-28) x AX146424 (1-4428)	

QY	1	AlaGlyProProAlaHisProArgProGluGluValGlyProProGluValAlaProGly	20
db	538	GCCTGCTCTCTCTGACCCAGCCGCCCTCTGGCACACTGTCGCATCTCTGGAGCCCTGGC	597
QY	21	LeuProGlnTyrThrGly	26
db	598	GCCTCCAGATATCCAGGT	615
RESULT 4	AC024805/c		
LOCUS	AC024805	68661 bp	DNA linear INV 29-MAY-2002
DEFINITION	Caenorhabditis elegans cosmid Y51H7C, complete sequence.		
ACCESSION	AC024805		
VERSION	AC024805.1	GI:7140366	
KEYWORDS	HTG.		
ORGANISM	Caenorhabditis elegans. Caenorhabditis elegans Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae; Rhabditidae; Pelodierinae; Caenorhabditis.		
REFERENCE	1 (bases 1 to 68661)		
AUTHORS	Waterston,R.		
TITLE	Genome sequence of the nematode C. elegans: a platform for investigating biology. The C. elegans Sequencing Consortium		
JOURNAL	Science 282 (5396), 2012-2018 (1998)		
PMID	99069613		
REFERENCE	9851916		
AUTHORS	2 (bases 1 to 68661)		
TITLE	Bradshaw-Cordum,H. and DuBague,T.		
JOURNAL	The sequence of C. elegans cosmid Y51H7C		
REFERENCE	Unpublished (2001)		
AUTHORS	3 (bases 1 to 68661)		
TITLE	Waterston,R.H.		
JOURNAL	Direct Submission		
REFERENCE	Submitted (01-MAR-2000) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA		
AUTHORS	4 (bases 1 to 68661)		
TITLE	Waterston,R.		
JOURNAL	Direct Submission		
REFERENCE	Submitted (24-MAR-2000) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA		
AUTHORS	5 (bases 1 to 68661)		
TITLE	Waterston,R.		
JOURNAL	Direct Submission		
REFERENCE	Submitted (24-APR-2001) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA		
AUTHORS	6 (bases 1 to 68661)		
TITLE	Waterston,R.		
JOURNAL	Direct Submission		
REFERENCE	Submitted (22-AUG-2001) Department of Genetics, Washington University, Genome Sequencing Center, 4444 Forest Park Avenue, St. Louis, MO 63110, USA		
AUTHORS	7 (bases 1 to 68661)		
TITLE	Waterston,R.		
JOURNAL	Direct Submission		
REFERENCE	Submitted (07-SEP-2001) Department of Genetics, Washington University, Genome Sequencing Center, 4444 Forest Park Avenue, St. Louis, MO 63110, USA		
AUTHORS	8 (bases 1 to 68661)		
TITLE	Waterston,R.		
JOURNAL	Direct Submission		
REFERENCE	Submitted (29-MAY-2002) Department of Genetics, Washington University, Genome Sequencing Center, 4444 Forest Park Avenue, St. Louis, MO 63110, USA		
AUTHORS	Submitted by:		
COMMENT	Genome Sequencing Center Department of Genetics, Washington University St. Louis , MO 63110, USA, and Sanger Centre, Hinxton Hall Cambridge CB10 1RO, England email: rwenematode.wustl.edu and jess@sanger.ac.uk		

NOTICE: This sequence may not be the entire insert of this clone.

It may be shorter because we only sequence overlapping sections once, or longer because we provide a small overlap between neighboring submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one m3 subclone.

For a graphical representation of this cosmid sequence and its analysis see:
<http://www.wormbase.org/db/seq/sequence?name=Y51H7C;class=Sequence>

NEIGHBORING COSMID INFORMATION

The 5' cosmid is F54D12, 200 bp overlap; the 3' cosmid is H17B01, 200 bp overlap.

NOTES:

coding sequences below are the result of integration and manual review of the following data : computer analysis using the program GeneFinder (P. Green and L. Hillier, personal communication), the large scale EST projects of Yuji Kohara (http://www.ddbj.nig.ac.jp/c-elegans/hm1/CE_INDEX.html) and The C. elegans ORFome cloning project (<http://worfdb.dcf.harvard.edu/>), similarity to other proteins from Blastx analyses (<http://blast.wustl.edu/>), sequence conservation with C. briggsae using Jim Kent's MABA alignment program (Genome Research 10:1115-1125, 2000), individual C. elegans Genbank submissions, and personal communications with C. elegans researchers. tRNAs are predicted using the program tRNAscan-SE (Lowe, T.M. and Dddy, S.R., 1997, Nucleic Acids. Res., 25, 955-964).

FEATURES

Source

```

/organism="Caenorhabditis elegans"
/strain="Bristol N2"
/db_xref="taxon:6239"
/chromosome="I"
/clone="Y51H7C"
585. .862
/gene="Y51H7C.8"
/note="for a graphical representation of this gene see:
http://www.wormbase.org/db/seq/sequence?name=Y51H7C.8
s=sequence"
join(585. .630,690. .862)

```

CDS

```

/gene-1217C.0
/codon_start=1
/product-"Hypochemical protein Y51HC.8"
/protein_id="AAK39332.1"
/db_xref="GI:13775524"
/translation="MSLIVQDELLGLCSQIQIRLNINVLCSIGKMKNEKGLRRVYV
VQSVVEIKVYKENGSGHQMDEDMELKf"
1328. 1834
/gene-"Y51HC.7"
/note-"for a graphical representation of this gene see:
http://www.wormbase.org/db/seq/sequence?name=Y51HC.7;class
s=Sequence"
join(1328..1450,1505..1630,1682..1834)
/gene="Y51HC.7"
/note="coded for by the following C. elegans cDNAs:
yk560h11.5, yk118a3.3, yk118a3.5, yk560h11.3"
/codon_start=1

```

CDS

```

/product-"Hypothetical protein Y51H7C.7"
/protein_id-"AAK39333.1"
/def_xref-"GI:13779525"
/translation-"MSSTRKSPADVIARLQKFEDELITLKEDAAVEEDVVDGKHPERSAA
HEALALVTSMFLMDSLMVAQATKGGAGRNDLLIDLPRTKRTADMKIKINLRQDA
PRNKQAAANFERNALTEQPOAGGESSKAAK"
2368..7728
/gene-"Y51H7C.6"

```

NOTICE: This sequence may not be the entire insert of this clone.

CDS

/note="for a graphical representation of this gene see:
http://www.wormbase.org/db/seq/sequence?name=y51H7C.6a;c1a
ss=Sequence"
join(2368..2797,2854..2979,3032..3261,4332..4496,
4570..5048,5864..6340,6802..7129,7558..7728)
/gene="y51H7C.6"

/note="coded for by the following C. elegans cDNAs:
yk4025.3, yk22a2.3, yk63f8.3, yk92f2.3, yk205b1.3,
yk56498.3, yk667a5.3, yk22a2.5, yk73d3.5, yk92f2.5,
yk15a3.5, yk184c3.5, CESSQ78F, yk205b1.5, yk399e12.5,
yk400b1.5, yk402e5.5, yk55592.3, yk55592.5, yk619e7.3,
yk667a5.5"

/codon_start=1
/product="Hypothetical protein y51H7C.6a"
/protein_id="AAK39338.1"
/db_xref="GI:13775530"

CDS

/translacion="MPEPILSHKSLGSLPAGRKNGVRHGEKPEKVGKQDFTRKIRE
LRLELEIKRREERIEKDIATILENTIDGGEONSFGIATYRLNNHMLVENSARKOL
TSALKNISVLADTISGRVSLADVAKTRVGCLOLADMDRGVCEGJIDDAIRSDFE
TASOHIFRFLTDOAVQIRREFKOKDADSTIRHSEVLSAKKERLSKILKSRLTSVQ
KGDVAEMORFIMRPLIHPREGLOKRVSPFNOKIDKLAENLAKMGKGDNDNRNV
LYADTLMEFEQVAETIESNLPVLEHSTGLKLDPMFTLQARIDEFRRLHEEDTR
RLSHFNRLVDVYIHOKKAAAEVADAPMEIDATASEICMMNTSAEMVWRFVSRIG
KNEYIRSPSGDDEENEAROEHRRLKREKAEKMDQLNRSRVGTMOELIGNYCL
LEHYMLKSVQKAIKSDYKEDAGLTSIIVDVVFIIKRSIRRAAGSNVDSVCATIN
NATALIDTVVGHILROSIOOGVTSNFSASFAFYAQQGPKVKEAADQKQEFLLALN
NSAKLSIELTILQGLITTEMAGVRPDKYKLESTQOJESAKKIASLKHGVEEL
FKSAFKKIKGADPQGITDROMQDVEYIEADHPFEQTLAQIDRLVNEPILFA
DNVGTLLTSSETARQIESLACQFNRYALQIDREYROIICATLVAVAGASAEKVK
GRIGQIYSLINVENTIDEAMEVHNSKTSATITRLTLPEVRKVALARAPFPVAIK
SIE"

join(2368..2797,2854..2979,3032..3261,4332..4496,
4570..5103)
/gene="y51H7C.6"
/note="coded for by the following C. elegans cDNAs:
yk400b1.3, yk400b1.5"

/codon_start=1
/product="Hypothetical protein y51H7C.6b"
/protein_id="AAK39339.1"
/db_xref="GI:13775531"

/translacion="MPEPILSHKSLGSLPAGRKNGVRHGEKPEKVGKQDFTRKIRE
LRLELEIKRREERIEKDIATILENTIDGGEONSFGIATYRLNNHMLVENSARKOL
TSALKNISVLADTISGRVSLADVAKTRVGCLOLADMDRGVCEGJIDDAIRSDFE
TASOHIFRFLTDOAVQIRREFKOKDADSTIRHSEVLSAKKERLSKILKSRLTSVQ
KGDVAEMORFIMRPLIHPREGLOKRVSPFNOKIDKLAENLAKMGKGDNDNRNV
LYADTLMEFEQVAETIESNLPVLEHSTGLKLDPMFTLQARIDEFRRLHEEDTR
RLSHFNRLVDVYIHOKKAAAEVADAPMEIDATASEICMMNTSAEMVWRFVSRIG
KNEYIRSPSGDDEENEAROEHRRLKREKAEKMDQLNRSRVGTMOELIGNYCL
LEHYMLKSVQKAIKSDYKEDAGLTSIIVDVVFIIKRSIRRAAGSNVDSVCATIN
NATALIDTVVGHILROSIOOGVTSNFSASFAFYAQQGPKVKEAADQKQEFLLALN
NSAKLSIELTILQGLITTEMAGVRPDKYKLESTQOJESAKKIASLKHGVEEL
FKSAFKKIKGADPQGITDROMQDVEYIEADHPFEQTLAQIDRLVNEPILFA
DNVGTLLTSSETARQIESLACQFNRYALQIDREYROIICATLVAVAGASAEKVK
GRIGQIYSLINVENTIDEAMEVHNSKTSATITRLTLPEVRKVALARAPFPVAIK
SIE"

CDS

/gene="y51H7C.9"
/note="for a graphical representation of this gene see:
http://www.wormbase.org/db/seq/sequence?name=y51H7C.9;c1a
ss=Sequence"
complement(join(7941..7992,8063..8169,8301..8422,
8469..8610,8657..8733,9106..9211,9619..9674..9817,
9865..9954,9999..10117,10177..10213))
/gene="y51H7C.9"
/note="coded for by the following C. elegans cDNAs:
yk362g7.3, yk362g7.5"
/codon_start=1
/product="Hypothetical protein y51H7C.9"
/protein_id="AAK39340.1"
/db_xref="GI:13775532"

Alignment Scores:

Pred. No.: 1.08e+03 Length: 68661
Score: 84.50 Matches: 16
Percent Similarity: 59.38% Conservative: 3
Best Local Similarity: 50.00% Mismatches: 8
Query Match: 52.16% Indels: 5
DB: 3 Gaps: 1

US-09-997-610-2_COPY_18_45 (1-28) x AC0204805 (1-68661)

.. QY 2 G1YPROPTALAHISPRGARGPRGROGLUGLVALG1YPROG1YALAPROGLYLeu 21

Db 66655 GGACCTCCAGGGGATTCCTGACCTCCGGGAGTACCTGGACCTCCGGGATTCCTGGACCA 66596

QY 22 PROGLTYR-----Thrglyguile 28

Db 66595 CCGGATACCGGTAGACCGGTACCTGGGCAATT 66560

RESULT 5

AC006799/C

LOCUS

DEFINITION

AC006799

AC006799

AC006799.1

AC006799.1

AC006799.1

AC006799.1

AC006799.1

AC006799.1

AC006799.1

AC006799.1

AC006799.1

AC006799.1

AC006799.1

AC006799.1

AC006799.1

AC006799.1

AC006799.1

AC006799.1

AC006799.1

AC006799.1

AC006799.1

AC006799.1

AC006799.1

AC006799.1

AC006799.1

AC006799.1

AC006799.1

AC006799.1

AC006799.1

AC006799.1

```

Db    66577   GGACCTCCAGG6GTTCCTGGACTCACC686GATCACTGCAGTCTCGTG6GTACTGSAACA    66518
Qy     22 ProIntTy-----ThretyGuIule 28
      ||::|||
      |::|::|
Db    66517   CCCGAATACCGGTAAAGAACCGTCACTG6GCCAATT        66482
RESULT 6
AX146430
DEFINITION Sequence 11 from Patent WO0134647.
ACCESSION AX146430
VERSION AX146430.1 GI:14284849
KEYWORDS
SOURCE
ORGANISM pig.
          Eumalvota; Metazoa; Chordata; Craniata; Vertebrata; Entelostomi;
          Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
REFERENCE 1 (bases 1 to 4428)
AUTHORS Bell,M.P., Neff,T.B., Polarek,J.W. and Seeley,T.W.
JOURNAL Animal collagens and gelatins
TITLE       Patent: WO 0134647-A 11 17-MAY-2001;
            FIBROGEN, INC. (US)
FEATURES
source location/Qualifiers
           1..4428 /organism="Sus scrofa"
BASE COUNT 994 a 1143 c 1381 g 910 t
ORIGIN
Alignment Scores:
Pred. NO.: 133 Length: 4428
Score: 84.00 Matches: 16
Percent Similarity: 61.54% Conservative: 0
Best Local Similarity: 61.54% Mismatches: 10
Query Match: 51.85% Indels: 0
DB: 6 Gaps: 0
US-09-997-610-2-COPY_18_45 (1-28) x AX146430 (1-4428)
Qy     1 AlaglyProProlahisProArpProprogluhuValGIYProProgIAAlAProGIly 20
      |||||||
      |||
Db    541 GCAGGTCCCCTGCGCCACCACTGTGCCCTTGTTACTGTGTACTCTGTGTGCCCCCTGTGT 600
Qy     21 LeuProGIInTyThrSIly 26
      |||
      |||
Db    601 TCTCCAGATACCAAGG 618
RESULT 7
AF017777 LOCUS AF017777 DEFINITION Drosophila melanogaster tweety (tly), flightless (fli), dodo (dod), penguin (pen), small optic lobes (sol), innocent bystander (lby), wacian (waw), bobby sox (bbx), sluggish (slg), helicase (hle), misto (mst), and la costa (lcs) genes, complete cds. ACCESSION AF017777 DB0043 U28044 VERSION AF017777.1 GI:3004652 SOURCE Drosophila melanogaster. ORGANISM Drosophila melanogaster. Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila. REFERENCE 1 (bases 1 to 66669) AUTHORS de Couet,H.G., Fong,K.S., Weeds,A.G., McLaughlin,P.J. and Miklos,G.L. TITLE Molecular and mutational analysis of a gelsolin-family member encoded by the flightless I gene of Drosophila melanogaster Genetics 141 (3), 1049-1059 (1995) MEDLINE PUBMED JOURNAL 8582612 REFERENCES 2 (bases 1 to 66669) Miklos,G.L., Yamamoto,M., Burns,R.G. and Malyszka,R. An essential cell division gene of Drosophila, absent from

```

JOURNAL	saccharomyces, encodes an unusual protein with tubulin-like and myosin-like peptide motifs
MEDLINE	Proc. Natl. Acad. Sci. U.S.A. 94 (10), 5189-5194 (1997)
PUBMED	97289742
REFERENCE	9144213
AUTHORS	3 (bases 1 to 6669)
TITLE	Maleszka,R., de Couet,H.G. and Mikiotis,G.L.
JOURNAL	Data transferability from model organisms to human beings: Insights from the functional genomics of the flightless region of Drosophila
MEDLINE	Proc. Natl. Acad. Sci. U.S.A. 95 (7), 3731-3736 (1998)
PUBMED	98188272
REFERENCE	9520435
AUTHORS	4 (bases 1 to 6669)
TITLE	Mikiotis,G.L.G., Kasprzak,A., Mason,J., de Couet,H.G., Hayward,D., Hall,D. and Maleszka,R.
JOURNAL	Direct Submission
COMMENT	Submitted (04-AUG-1997) Visual Sciences, Research School of Biological Sciences, The Australian National University, Sullivan Creek Rd., Canberra, ACT 0200, Australia On or before Apr 2, 1998 this sequence version replaced gi:1052878. gi:1052878.
FEATURES	Location/Qualifiers
SOURCE	1..6669
gene	/organism="Drosophila melanogaster" /strain="CantonS" /db_xref="taxon:7227" /chromosome="X" complement(1639..7694) /gene="tly"
mRNA	complement(join(1639..2707,2789..2800,3622..3791, 3917..4680,4745..5230,5525..5809,7521..7694)) /gene="tly"
CDS	/product="tweety" complement(join(1767..2707,2789..2800,3622..3791, 3917..4680,4745..5230,5525..5662)) /gene="tly" /codon_start=1 /product="tweety" /protein_id="AAC28400.1" /db_xref="GI:3004653" /translation="MGDIHEFTDYKVPYIAKLHALPHNYTFHKINSTFRNDITY LESIGLGSVPAAALLVSLGLGLEYLMTRCCDRKPRAMISTLSLKVALSTVTWCGAA IGIGIVSGNDIDHNGLEIVETAGRKVDNLVTITRNDHTLENTLNTRIRPLVELAIF DQPSNQTALSKLFESLVIVQGAVTLTATASDIRPLMISMTHELTRDDOMELRM PGVATFALLVLCAVLLGVARSHACGLEFSVGILAVYSMLSGILSSVAVG DLCSPLADPVSATAPRDLPNVLTLYHTCOEELFHNPETFORRESONSINARSAMTV MKISVIFKSSGGIQRKIGAVNADINSERBLTOLFALVDCAYAHNLAARLCCEGG LLGLVMIALISFIALILLTLMWVDSHWITVRKNDYAQVDESYTISHAPQNHOOM MNAARTLPKRHHNGSESPVITSGSHTLOHPKROQHKNMAHAHIQQNRKAGTHTLRRL PSHNSTPTMHGTGNANAAVAAAAMAANNPPTQAQQOQQOQQOQQOQQOQLGSP OPITCRHPHDPHPHPHOSHSAVAAVVOHAIHYHQOQAOQVGYTTTAHHAP HLHGPGOSIOYICOTIPAHILAPOALAANGNPISYCPRIYVSGSIYVSMATMRSONSG GPQIPRAOHPPSLHQOQQOQPPEPSQOOQOLHOKSPQOHQOQLOOHOQOQQOHHQOQ HHQOQHOOQOHHQOQQOQPNESDVVEITSTADOTATYDKDKQIKYCSTLRGGGFEDPKY KPLINCPIPEIKPDKEQRPVESIVEORQAHHQVYSK" complement(15809..7521) /gene="tly" 8586..13783 /gene="f11" join(8586..8705,8669..12023,12260..12410,12533..13783) /gene="f11" /product="flightless" join(8652..8705,8869..12023,12260..12410,12533..12943) /gene="f11" /codon_start=1 /product="flightless" /protein_id="AAC28407.1" /db_xref="GI:3004660" /translation="NSVIFPVFGVDFETKNDTSATPPSQBMQSRVQITIDTPQALAI PEEGIQAKTEHLSTLNNRLKRTGEETIELSCSLSDRHQNLKNSITPEELFEELE TTLDLSNNKLKEVPEGERAKNLIIVNLISNQDISIPPLETILHTLDLFLDSHNLE TLTPQTRNLTKTIDLSHNPLEQLDQLSPQLSDSELNETEITAGVETIMORTESLNSNO ANICEIDLNSNKLPKPCCVYVVWVTLVRLNISDNELNETAGVETIMORTESLNSNO
CDS	

* 36719 36818: gap of unknown length
* 36819 36835: contig of 153 bp in length
* 38356 38455: gap of unknown length
* 38456 39544: contig of 1089 bp in length
* 39545 39644: gap of unknown length
* 39645 42097 42096: contig of 2453 bp in length
* 42097 42196: gap of unknown length
* 42197 44438: contig of 2242 bp in length
* 44439 44538: gap of unknown length
* 44539 46550: contig of 2112 bp in length
* 46551 46751 46750: gap of unknown length
* 46751 48937 48936: contig of 2186 bp in length
* 48937 49035: gap of unknown length
* 49037 52140: contig of 5104 bp in length
* 52141 52240: gap of unknown length
* 52241 54747: contig of 2507 bp in length
* 54748 54848 54847: gap of unknown length
* 54848 59018 59017: contig of 4170 bp in length
* 59018 59117: gap of unknown length
* 59118 62801: contig of 3684 bp in length
* 62801 62902 62901: gap of unknown length
* 62902 66517: contig of 3616 bp in length
* 66518 70274 70273: contig of 3657 bp in length
* 70275 70375 70374: gap of unknown length
* 70375 72539: contig of 5165 bp in length
* 72539 75340 75339: gap of unknown length
* 75340 79309: contig of 3670 bp in length
* 79310 79409: gap of unknown length
* 79410 84689: contig of 5280 bp in length
* 84690 84789: gap of unknown length
* 84790 88910 88910: contig of 5021 bp in length
* 88911 89910 89910: gap of unknown length
* 89911 95503: contig of 5593 bp in length
* 95504 95604 95603: gap of unknown length
* 100929 101028 101028: contig of 5325 bp in length
* 101029 105513: gap of unknown length
* 105514 105613: gap of unknown length
* 105614 110389: contig of 4776 bp in length
* 110390 110489: gap of unknown length
* 110490 114853: contig of 4363 bp in length
* 114853 114953 114952: gap of unknown length
* 114953 120725: contig of 5774 bp in length
* 120727 120826: gap of unknown length
* 120827 127059: contig of 6233 bp in length
* 127060 133607: gap of unknown length
* 133608 133707: contig of 6448 bp in length
* 133708 139628: gap of unknown length
* 139629 139728: contig of 5921 bp in length
* 139729 147785: gap of unknown length
* 147786 147885: gap of unknown length
* 147886 154861: contig of 6976 bp in length
* 154862 154961: gap of unknown length
* 154962 160982: contig of 6021 bp in length
* 160983 161082: gap of unknown length
* 161083 167776: contig of 6694 bp in length
* 167777 167876: gap of unknown length
* 167877 175123: contig of 7247 bp in length
* 175124 175224 175223: gap of unknown length
* 175224 183464: contig of 8241 bp in length
* 183465 183564: gap of unknown length
* 183565 191380: contig of 7816 bp in length
* 191381 191480: gap of unknown length
* 191481 200933: contig of 9453 bp in length
* 200933 201033: gap of unknown length
* 201033 207587: contig of 6355 bp in length
* 207588 207687: gap of unknown length
* 207688 216967: contig of 9280 bp in length
* 216968 229646: gap of unknown length
* 229647 229746: gap of unknown length

* 229747 241492: contig of 11746 bp in length
* 241493 241592: gap of unknown length
* 241593 256417: contig of 14825 bp in length.
Location/Qualifiers
source 1..256417
/organism="Mus musculus"
/db_xref="taxon:10090"
/chromosome="16"
/clone="RP23-109A22"
/clone_lib="RPCI mouse BAC library 23"
BASE COUNT 67041 a 56625 c 61089 g 63261 t 8401 others
ORIGIN
Alignment Scores:
Pred. No.: 3.44e+03 Length: 256417
Score: 84.00 Matches: 15
Percent Similarity: 68.18% Conservative: 0
Best Local Similarity: 68.18% Mismatches: 7
Query Match: 51.85% Indels: 0
DB: 2 Gaps: 0
US-09-997-610-2_COPY_18_45 (1-28) x AC079422 (1-256417)
Qy 4 ProAlaHisProArgProProGluGluValGlyProProGlyAlaProGlyLeuProGln 23
Db 100552 CCGCGCCCCCTGCCCCCGACCCCGCGCCCCCGCCCCCGCCCCGCTNCCCA 100611
Qy 24 TyrThr 25
Db 100612 ACNACG 100617
RESULT 12
AE003568 292911 bp DNA linear INV 05-OCT-2000
LOCUS Drosophila melanogaster genomic scaffold 142000013386037 section 5
DEFINITION of 5, complete sequence.
ACCESSION AE003568 AE002620
VERSION AE003568.2 GI:10726967
KEYWORDS HTG.
SOURCE Drosophila melanogaster.
ORGANISM Drosophila melanogaster.
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
REFERENCE
1 (bases 1 to 292911)
Adams,M.D., Celisier,S.E., Holt,R.A., Evans,C.A., Gocayne,J.D.,
Amanatides,P.G., Scherer,S.E., Li,P.W., Hoskins,R.A., Galie,R.F.,
George,R.A., Lewis,S.E., Richards,S., Ashburner,M., Henderson,S.N.,
Sutton,G.G., Wortman,J.R., Vandeil,M.D., Zhang,Q., Chen,L.X.,
Brandon,R.C., Rogers,Y.H., Blazey,R.G., Champs,M., Pfeiffer,B.D.,
Wan,K.H., Doyle,C., Baxter,E.G., Helt,G., Nelson,C.R., Gabor
Miklos,G.L., Abril,J.F., Agbayani,A., An,H.J.,
Andrews-Pfankoch,C., Baldwin,D., Bailew,R.M., Basu,A.,
Baxendale,J., Bayraktaroglu,L., Beasley,E.M., Beeson,K.Y.,
Bens,P.V., Berman,B.P., Bhandari,D., Bolshakov,S., Borkova,D.,
Botchan,M.R., Bouck,J., Brokstein,P., Brothier,P., Butts,K.C.,
Busam,D.A., Butler,H., Cadieu,E., Center,A., Chandra,T.,
Cherry,J.M., Cawley,S., Dahlke,C., Davenport,L.B., Davies,P., de
Pablos,B., Delcher,A., Deng,Z., Mays,A.D., Dew,I., Dietz,S.M.,
Dodson,K., Doup,L.E., Downes,M., Dugan-Rocha,S., Dunkov,B.C.,
Dunn,P., Durbin,K.J., Evangelista,C.C., Ferraz,C., Ferreira,S.,
Fleischmann,W., Foster,C., Gabriellian,A.E., Garg,N.S.,
Gelbart,W.M., Glasser,K., Glodek,A., Gong,F., Gottrell,J.H., Gu,Z.,
Guan,P., Harris,M., Harris,N.L., Harvey,D., Helman,T.J.,
Hernandez,J.R., Houck,J., Hostin,D., Houston,K.A., Howland,T.J.,
Wei,M.H., Ibegwam,C., Jalali,M., Kalush,F., Karpen,G.H., Ke,Z.,
Kennison,D.A., Ketchum,K.A., Kimmel,B.E., Kodira,C.D., Kraft,C.,
Kravitz,S., Kulp,D., Lai,Z., Lasko,P., Lei,Y., Levitsky,A.A.,
Li,J., Li,Z., Liang,Y., Lin,X., Liu,X., Mattei,B., McIntosh,T.C.,
McLeod,M.P., McPherson,D., Merkulov,G., Milshina,N.V., Moharry,C.,
Morris,J., Mosher,J., Mount,S.M., Moy,M., Murphy,B., Murphy,L.,
Muzny,D.M., Nelson,D.L., Nelson,D.R., Nelson,K.A., Nixon,K.,
Nusskern,D.R., Pacleb,J.M., Palazzolo,M., Piltman,G.S., Pan,S.,

Pollard,J., Puri,V., Reese,M.G., Reimert,K., Remington,K.,
Saunders,R.D., Scheeler,F., Shen,H., Shue,B.C., Siden-Kiamos,I.,
Simpson,M., Skupski,M.P., Smith,T., Spier,E., Spradling,A.C.,
Stapleton,M., Strong,R., Sun,E., Swirskas,R., Tector,C., Turner,R.,
Venter,E., Wang,A.H., Wang,X., Wang,Z.Y., Wasserman,D.A.,
Weinstock,G.M., Weissenbach,J., Williams,S.M., Woodage,T.,
Worley,K.C., Wu,D., Yang,S., Yao,Q.A., Ye,J., Yeh,R.F.,
Zaveri,J.S., Zhan,M., Zhang,G., Zhao,Q., Zheng,L., Zheng,X.H.,
Zhong,F.N., Zhong,W., Zhou,X., Zhu,S., Zhu,X., Smith,H.O.,
Gibbs,R.A., Myers,E.W., Rubin,G.M. and Venter,J.C.
The genome sequence of *Drosophila melanogaster*
Science 287 (5461), 2185-2195 (2000)

JOURNAL MEDLINE
PUBMED 10731132
2 (bases 1 to 292911)

REFERENCE
ADAMS,M.D., Celniker,S.E., Gibbs,R.A., Rubin,G.M. and Venter,C.J.
Direct Submission
Submitted (21-MAR-2000) Celera Genomics, 45 West Gude Drive,
Rockville, MD, USA

COMMENT
On Oct 9, 2000 this sequence version replaced gi:7295493.

FEATURES
source
Location/Qualifiers
1..292911
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/chromosome="X"
<8329..>8511
/gene="CG14580"
/evidence=not_experimental
/map="19F1-19F1"
/db_xref="FLYBASE:FBgn0014580"
/db_xref="FLYBASE:FBgn0040646"
<8329..>8511
/gene="CG14580"
/product="CT34315"
/evidence=not_experimental
/db_xref="FLYBASE:FBgn0014580"
/db_xref="FLYBASE:FBgn0040646"
8329..8511
/gene="CG14580"
/note="CG14580 gene product"
/codon_start=1
/evidence=not_experimental
/protein_id="AAF50841.1"
/db_xref="GI:7295528"
/db_xref="FLYBASE:FBgn0014580"
/db_xref="FLYBASE:FBgn0040646"
/translation="MKDVSWEQKEHKEQKQEQEQKVEDQKSTAFISNQLPPS
GNSINPPIPTLNATF"
complement(22044..>23009)
/gene="CG1722"
/map="19F1-19F1"
/db_xref="FLYBASE:FBan0001722"
/db_xref="FLYBASE:FBgn0031168"
complement(22044..>23009)
/gene="CG1722"
/product="CT4924"
/db_xref="FLYBASE:FBan0001722"
/db_xref="FLYBASE:FBgn0031168"
complement(22044..>22886)
/gene="CG1722"
/note="CG1722 gene product"
/codon_start=1
/protein_id="AAF50840.1"
/db_xref="GI:7295527"
/db_xref="FLYBASE:FBan0001722"
/db_xref="FLYBASE:FBgn0031168"
/translation="MPKRSRSPKATKPLSTKRLQSKVKSQKTKPKMOOKSVKOP
LKKPMRSPIMRAALOPTILGAKKPKKSVTKKSHRPLDAMQSLGMSRRPLI
DASPKRKAQTSVLELGRDSRTTKPKKPKSPSEKPKSKNOLPKLRMERVMTKA
ETRAIDGVNFCNLGVKMQPVTKRYQQLQKGEAGFEGRLAGEGSTRGDALRRPV
RTQPSGLSLVKGAVIGAFADVIVPLGRPLIIDPISTKEFQNMNSAASKKPKPSHK
NTHL"
<27150..>27446
gene

/gene="CG14581"
/evidence=not_experimental
/map="19F1-19F1"
/db_xref="FLYBASE:FBan0014581"
/db_xref="FLYBASE:FBgn0040645"
<27150..>27446
/gene="CG14581"
/product="CT34316"
/evidence=not_experimental
/db_xref="FLYBASE:FBan0014581"
/db_xref="FLYBASE:FBgn0040645"
27150..27446
/gene="CG14581"
/note="CG14581 gene product"
/codon_start=1
/evidence=not_experimental
/protein_id="AAF50839.1"
/db_xref="GI:7295526"
/db_xref="FLYBASE:FBan0014581"
/db_xref="FLYBASE:FBgn0040645"
/translation="NYHMGSGKKRKSQMLDIYSREKFAPIQETILRIELEDL
NSGRQSLPSIETDIQTARLNQKLEEMVITRMGPTLQADNIIENAVEVSS"
<38623..>63995
/gene="CG1494"
/map="19F1-19F2"
/db_xref="FLYBASE:FBan001494"
/db_xref="FLYBASE:FBgn0031169"
join(58623..59915,59983..60362,60585..61037,61265..61605,
61758..62055,62994..63743,63815..>63995)
/gene="CG1494"
/product="CT3739"
/db_xref="FLYBASE:FBan001494"
/db_xref="FLYBASE:FBgn0031169"
join(58725..59915,59983..60362,60585..61037,61265..61605,
61758..62055,62994..63743,63815..>63995)
/gene="CG1494"
/note="CG1494 gene product"
/codon_start=1
/protein_id="AAF50838.1"
/db_xref="GI:7295525"
/db_xref="FLYBASE:FBan001494"
/db_xref="FLYBASE:FBgn0031169"
/translation="WVAFFRNALNARVLTILMISYVPTFLSNLGNIALRY
SYALPNVATLVEIEFLIERESIHTWEDSGRYLNDGGHITVTAASMJEMALVYC
AIGLYDMRGDGRSGKKKPKPTNANSQEDPYHEGDSFTHGOAIGVNSKIYVE
PSHRRKTLTKLCRFATNDPRLNLFNSVNYENETVLMGNGCKSTLKTILGL
VESRRCTVMISSHNIOETERRKAAMELGIAFGHMLTGFVIDYLFICRVGLHNNI
EIDGSNYELNVLQIGLTKTARIETLDDDLVASCFAFVGNSTYLLIDVHSDLDK
RQOSLYNLIINEKSKRTIILVNSPALENIDRMAINSNGELCKTGKPKLKNYG
HCYRLTCVGKNTKRDLEFGMNSVPMNMSIERDIGKYKTVLENKEEDQPMILIDL
EEMQOLGVYSFRIRDTSMETFLFRGCEDNDSQAFOSHENAQVLLERYSTLAEN
EKGRRGKLFELHGRAVLYKRWIAAHRIWLVLEFVLAALVACTESSIFLYGKNY
ELEPLTFNLSQLHTVDAFVLESEEDVDMAHYVELLWDAHAATLKNHNAVA
LITONQFTHVNSRYIFGATFDKITYAFNNNPLLSAPYALVAVNAVARHLFDEEA
TIDVTAPLPPRTAINTPPSSHTPGCGAFCGCVLTFEIPMAFAYIMTERGSLTK
OOFIAGVRCYWTFTVYLDLLEFLIFCGAYGVNVALYENPNHDAVLGYISTLMIG
GFWIILATLMAISLCRNPCTGFLMLGINSIGLVCQSQRYRTHPSZMLLEPTMAIT
VAVTICKLFMTIYEKFLICMDPVNFTSVSEKSLITGANSKVTTLKLVANETK
MNAGOLMIHIDVSYNTHRVOCYRWAGYCPKDSIPSEFTREGLNYTHAMLGHRHGR
ELSEALLRLVGLPCWNRSVRMCTGOIRLRYFAVAVLSQPLMIDCVAGPAGIDPGK
RIIIMTSTMOAGSSELYTMLTGDAERLSLPTLLGOLMIMIDVPMPTENTENYSG
YOLEVFRKKVNPVNSRATWNLINHPMSPKKRSAPMEIKFPAVYLTIRBDSNV
FVPLCTTFTESEFLTLRKDAPEMNIEDYFITRNMLVGFQIFTYDOHDP"
complement(64564..>72212)
/gene="CG1718"
/map="19F1-19F2"
/db_xref="FLYBASE:FBan001718"
/db_xref="FLYBASE:FBgn0031170"
complement(join(64564..65380,65443..66216,66361..66537,
66595..66745,66950..67802,67867..68234,68300..70124,
70196..70784,72001..72212))
/gene="CG1718"
/product="CT4892"
mRNA
gene

Thu Feb 20 14:55:13 2003

us-09-997-610-2_copy_18_45.rge

Page 13

Search completed: February 19, 2003, 23:10:55
Job time : 581.136 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: February 19, 2003, 14:53:22 ; Search time 1025.49 seconds

(without alignments)
2084.671 Million cell updates/sec

Title: US-09-997-610-2_COPY_18_149

Perfect score: 712

Sequence: 1 AGRNHRPRPREVGPARGA.....KEISKQSQSIQEVTVLKA 132

Scoring table: BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 08
Maximum Match 1008
Listing first 45 summaries

Command line parameters:
-MODEL=frame_plus_p2n.model -DEV=tlh
-Q/cgn2.1/USPTO_SPOOL/US09997610/runatc.10022003-160823-23843/app_query.fasta.1.1635
-DB=EST -QFMT=fastap -SUFFIX=1st -MINMATCH=0.1 -LOOPEXT=0 -LIST=45
-UNITS=bits -START=1 -END=1 -MATRIX=bloms62 -TRANS=human40.ccl -ALIGN=45
-DOCALLIGN=200 -THR.SCORE=DCT -THR.MAX=100 -THR.MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEADSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US09997610.ecgn.1.1.1871 -runatc.10022003-160823-23843 -NCPU=6 -ICPU=3
-NO_XLPPX -NO_MMAP -LARGEORDER -NEG.SCORES=0 -WAIT -LONGLOG -DEV.TIMEOUT=120
-XARN.TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -Fgapop=6 -Fgapext=7
-YGAPOP=10 -YGAPEXT=0.5 -DELPEXT=7

Database :
1: em_estba:*
2: em_estbhm:*
3: em_estcin:*
4: em_estcinu:*
5: em_estcov:*
6: em_estcpl:*
7: em_estcro:*
8: em_hlc:*
9: gb_estc1:*
10: gb_estc2:*
11: gb_hlc:*
12: gb_est3:*
13: gb_est4:*
14: gb_estc5:*
15: em_estfun:*
16: em_estfun:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pin:*
21: em_gss_vit:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_tod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description
1	306.5	43.0	499	14	BM967732
2	228.5	32.1	526	12	BE757276
3	226.5	31.8	504	12	BE757275
4	181	25.4	663	13	BJ068972
5	165.5	23.2	1092	17	CNS0540B
6	158.5	22.3	512	13	BJ494317
7	158	22.2	587	10	AV593019
8	157.5	22.1	689	10	BB318780
9	157.5	22.1	2462	11	AK018742
10	157	22.1	662	14	BQ419957
11	155.5	21.8	695	9	AV243093
12	154.5	21.7	921	9	AA673154
13	152.5	21.4	468	10	AW988554
14	152.5	21.4	532	10	BE625509
15	152.5	21.4	633	10	BB625284
16	152.5	21.4	707	12	BC872725
17	152.5	21.4	921	11	AK003138
18	151.5	21.3	460	9	A1332213
19	150.5	21.1	508	9	AA493090
20	150.5	21.1	589	14	BQ193813
21	150.5	21.1	885	9	A1323115
22	149	20.9	589	13	B1760895
23	149	20.9	942	14	BQ900894
24	147.5	20.7	856	14	BQ876848
25	146.5	20.6	607	12	BC691150
26	146	20.5	585	10	AW913280
27	145.5	20.4	580	12	BC691162
28	145.5	20.4	1029	17	CNS022P5
29	145.5	20.4	453	10	BE126709
30	144.5	20.3	639	13	BJ090235
31	144.5	20.3	430	10	AA453561
32	143	20.1	633	9	AA060543
33	142.5	20.0	890	14	BQ737531
34	141.5	19.9	683	10	AM029321
35	141.5	19.8	529	12	BF074104
36	140.5	19.7	482	14	BQ260214
37	140	19.7	542	14	BQ449231
38	140	19.7	563	14	BQ285664
39	140	19.7	566	14	BQ285785
40	140	19.7	580	13	BI477496
41	140	19.7	587	14	BQ420027
42	140	19.7	587	14	BQ285786
43	140	19.7	598	14	BQ420183
44	140	19.7	626	14	BQ260132
45	140	19.7			

ALIGNMENTS

RESULT 1
LOCUS BM967732
DEFINITION LM24HM0134 Bos taurus LM-24-HW CDNA library Bos taurus cDNA clone
ACCESSION BM967732
VERSION BM967732.1 GI:19561919
KEYWORDS EST.
SOURCE COW.
ORGANISM Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Rumiantia; Pecora; Bovidae;
Bovinae; Bos.
REFERENCE 1 (bases 1 to 499)
Yoon,D.H., Jang,Y.S., Kim,T.H., Park,E.W., Lee,H.K., Chung,E.R.,

TITLE Sun, S. S. and Cheong, I. C.
JOURNAL Unpublished (2002)
COMMENT Contact: Dr. Du-Hak Yoon
National Livestock Research Institute, RDA
564 Omockcho-dong, Suwon, 441-350, Korea
Tel: 82 31 290 1593
Fax: 82 31 290 1792
Email: dhyoon@rda.go.kr
Insert Length: 499 Std Error: 0.00
Seq primer: CAGGAACAGCTATGAC
POLYA-NO.

FEATURES location/Qualifiers
source 1..499
/organism="Bos taurus"
/db_xref="taxon:9913"
/clone_lib="LM-24-HW-011-34 (5')"
/clone_lib="Bos taurus LM-24-HW cDNA library"
/sex="Six males mixed"
/tissue_type="longissimus dorsi"
/cell_type="myocyte"
/dev_stage="24 months old"
/lab_host="XLI-BlueMRP strain"
/note="Organ: skeletal muscle; Vector: Uni-ZAPXR; Site_1:
EcORI: Site_2: Xho I"

BASE COUNT 129 a 139 c 131 g 100 t

ORIGIN

Alignment Scores:
Pred. No.: 2,55e-23 Length: 499
Score: 306.50 Matches: 63
Percent Similarity: 67.62% Conservative: 8
Best Local Similarity: 60.00% Mismatches: 28
Query Match: 43.05% Indels: 6
Gaps: 2

US-09-997-610-2_COPY_18_149 (1-132) x BM967732 (1-499)

OY 2 GtYpRoPtoAlaHnIsPrOArGrProGlUglUValGlyProPro-----GlyAla 18
|||||
Db 183 GGGCCACGACGACCCAGGCGCTCTCAAGATTAAGAGGCGCTCTCAAGAGGAATA 242
|||||
OY 19 ProGlyLeuProGlnTyrThrGlyGluLeSerGluMetThrIlyscysProGlyscPrasp 38
|||||
Db 243 CCAAGGTTGGCCAGGTCCCGCGGAATCCAGAGCAAGTGTAAAGCCCATGACAC--- 299
|||||
OY 39 lIeGlUArGSeRAlaPheThValIySLeuSeRcLyLySLeuProLeuProPheIySpro 58
:::|||||
Db 300 ---AGACAGTCGCGCTTACTGTGACGTGACGTGCGCAGTTCGCCCTTCAAAAGCCT 356
|||||
OY 59 lIe-lIepheThrGlyValLeuTyTAsnAlaGlnArGAspLeuLySgLuAlaMetGlyVa 78
:::|||||
Db 357 GTGCCCTTTCACAGAGGCTGTGTAACAATTCCTTTCTATGTGATTCCTCATCTG 416
|||||
OY 78 lPheAlaCySaRgValPProGlyAsnTyTyTyrSerSerPheAspValGlnLeuHIScY 98
|||||
Db 417 GTTCACATGACAGGTGCGCAGGAATTCACATTCCTTTCTATGTGATTCCTCATCTG 476
|||||
OY 98 slySValAsnIle 102
|||||
Db 477 CAAGGTGACTGTT 489

RESULT 2
BE757276 526 bp mRNA linear EST 25-APR-2001
LOCUS BE757276
DEFINITION 211669 MARC 2BOV Bos taurus cDNA 5', mRNA sequence.
ACCESSION BE757276
VERSION BE757276.1 GI:10171268
KEYWORDS EST.
SOURCE Bos taurus
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;

REFERENCE Bovidae; Bovinae; Bos.
AUTHORS 1 (bases 1 to 526)
Smith, T.P.L., Grosse, W.M., Freking, B.A., Roberts, A.J., Stone, R.T.,
Casas, E., Wray, J.E., White, J., Cho, J., Fahrenkrug, S.C., Bennett,
G.L., Heaton, M.P., Laegreid, W.W., Rohrer, G.A., Chitko-Mckown, C.G.,
Perteira, G., Holt, I., Karamycheva, S., Liang, F., Quackenbush, J., and
Keele, J.W.
TITLE Sequence evaluation of four pooled-tissue normalized bovine cDNA
JOURNAL libraries and construction of a gene index for cattle
MEDLINE Genome Res. 11 (4), 626-630 (2001)
COMMENT Contact: Smith TPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390
Email: smith@email.marc.usda.gov
Single pass sequencing. Bases called and alt-trimmed with phred
v0.980904.e. Vector identified by cross-match with the -minscore 18
and -minmatch 12 options.
PCR Primers
FORWARD: AGGAACAGCTATGACCAT
BACKWARD: GTTTCACATGACGACG
Plate: 63 row: G column: 14
Seq primer: ATTAGGTGACACTATGAC.

FEATURES location/Qualifiers
source 1..526
/organism="Bos taurus"
/db_xref="taxon:9913"
/clone_lib="MARC 2BOV"
/tissue_type="pooled"
/lab_host="DH10B"
/note="Vector: pCMV SPORT6; Site_1: NotI; Site_2: SalI;
Library made from pooled tissue from testis, thymus,
semiteudonous muscle, longissimus muscle, pancreas,
adrenal, and endometrium"

BASE COUNT 106 a 137 c 153 g 130 t

ORIGIN

Alignment Scores:
Pred. No.: 8,06e-15 Length: 526
Score: 228.50 Matches: 52
Percent Similarity: 55.26% Conservative: 11
Best Local Similarity: 45.61% Mismatches: 40
Query Match: 32.09% Indels: 11
Gaps: 3

US-09-997-610-2_COPY_18_149 (1-132) x BE757276 (1-526)

OY 2 GtYpRoPro-----AlaHIsPrOArGrProGlUglUValGlyProProGlyAla 18
|||||
Db 181 GGACCTCCCGGACTCAGTGTCTTCAGAGGCGACCTGTTATGACAGCCCATAGGTATG 240
|||||
OY 19 ProGlyLeu-----ProGlnTyrThrGlyGluLeSerGluMetThrIlyscysPro 35
|||||
Db 241 CCAAGGATACAGGAGACCTGGGCTTCCTGACTGCTTAGAGAG----- 285
|||||
OY 36 CySPRoAspIlEglUArGSeRAlaPheThValIySLeuSeRcLyLySLeuProLeuPro 55
|||||
Db 286 TGCCACCCCTGCTCCTACGTTCGCTTTTCCTCAAGTGAAGTGGGCTTTCACAGAGACC 345
|||||
OY 56 PheLySPRoIlEilEphThrGlyValLeuTyTAsnAlaGlnArGAspLeuLySgLuAla 75
:::|||||
Db 346 TCCAGCCCATGTCTTCACAGGAAGTTCTGTACAAACCATAGGCGCCACTTGCAGCCCGCC 405
|||||
OY 76 MetGlyValPheAlaCySaRgValPProGlyAsnTyTyTyrSerSerPheAspValGlnLeu 95
|||||
Db 406 ACTGtGtGtTtAGCTGACGAGGCTGCTGtGtGtTtACCACTTtGgCTTtGACtTtGAGtTtG 465
|||||
OY 96 HIsHIScYsLySValAsnIleTyrPLeuMetArGlySgLnIle 109
|||||
Db 466 TTTCAAGTGTCTGACAGGTGGGTCTATGTGCGGAATGCGATC 507

```

RESULT 3
BE757275          504 bp  mRNA  linear  EST 25-APR-2001
LOCUS             BE757275
DEFINITION        211668 MARC 2BOV Bos taurus cDNA 5', mRNA sequence.
ACCESSION         BE757275
VERSION           BE757275.1  GI:10171267
KEYWORDS          EST.
SOURCE            cow.
  ORGANISM        Bos taurus
    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
    Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
    Bovidae; Bovinae; Bos.
  REFERENCE       1 Smith,T.P.L., Grosse,W.M., Freking,B.A., Roberts,A.J., Stone,R.T.,
    AUTHORS        Casas,E., Wray,J.E., White,J., Cho,J., Fahrnkung,S.C., Bennett
    ,G.L., Heaton,M.P., Laegreid,W.W., Rohrer,G.A., Chitko-McKown,C.G.,
    ,Pertea,G., Holt,I., Karamychewa,S., Liang,F., Quackenbush,J. and
    Keefe,J.W.
    TITLE          Sequence evaluation of four pooled-tissue normalized bovine cDNA
    JOURNAL        libraries and construction of a gene index for cattle
    MEDLINE        Genome Res. 11 (4), 626-630 (2001)
    COMMENT        21180013
    CONTACT: Smith TPL
    USDA, ARS, US Meat Animal Research Center
    PO Box 166, Clay Center, NE 68933-0166, USA
    Tel: 402 762 4366
    Fax: 402 762 4390
    Email: smithhemall.marc.usda.gov
    Single pass sequencing. Bases called and alt-trimmed with phred
    v0.980904.e. Vector identified by cross_match with the -linscore 18
    and -minmatch 12 options.
    PCR primers
    FORWARD: AGGAACGACGTATGACCAT
    BACKWARD: GTTTTCCAGTCACGACG
    Plate: 63 row: 6 column: 13
    Seq primer: ATTTAGTGACACATATAG.
FEATURES
  source          1..504
                  /organism="Bos taurus"
                  /db_xref="taxon:9913"
                  /clone_1id="MARC 2BOV"
                  /tissue_type="Pooled"
                  /lab_host="DH10B"
                  /note="Vector: pCMV SPORT6; Site_1: NotI; Site_2: SalI;
                  Library made from pooled tissue from testis, thymus,
                  semitendinosus muscle, longissimus muscle, pancreas,
                  adrenal, and endometrium."
BASE COUNT       99 a 131 c 146 g 128 t
ORIGIN
Alignment Scores:
Pred. No.:       1,25e-14      Length: 504
Score:           226.50        Matches: 51
Percent Similarity: 55.86%     Conservative: 11
Best Local Similarity: 45.95%  Mismatches: 38
Query Match:     31.81%       Indels: 11
DB:              12           Gaps: 3
US-09-997-610-2_copy_18_149 (1-132) x BE757275 (1-504)
QY 2 GIYProPro-----AlAhISProARgProProGIUGlUaIGlYProProGIaYAla 18
DB 161 GGAACCTCCGGGACGACGTGCTCTCCAGGCGCACCTGGTTATACAGACCCATAGGTATG 240
QY 19 ProGIaUa-----ProGIaUaThrGIaUaIleSerGIaUaMetThrIleScysPro 35
DB 241 CCAGAGACTAACAGGAGGAGACCTGGCTCTCGTACGTCGTGTAAGAG-----285
QY 36 CysProAspIleGIaUaSerAlaPheThrValIleSleuSerGIaUaProleuPro 55
DB 286 TGCACACCCCTGCTCAGTCGCTTTTCCGTCAGAGCTGAGGGCCCTTTCCACAGAGCC 345
QY 56 PheUysProIleIlePheThrGIaUaIleuTyraAsnAlaGIaUaArgAspIleuSglUa 75

```

```

DB 346 TCCACAGCCATGTGTCTTCAGAGAAATTCTGTACAAACATCAGGCGCCATTCAGACCCCGCC 405
QY 76 MetGIaUaIleAlaCysArgValProGIaUaThrTyrTyrSerSerPheAspValGIaUa 95
DB 406 ACTGGTGTCTGACGTCAGAGCGTCCTGGTGTGTACACATTGGCGTTGACATTGAGATTG 465
QY 96 HlSHIScYsLysValAsnIleThrPleuMetArg 106
DB 466 TTTTCAGAGTGTCTGTCAGGTGGGTCTTAATGCGG 498
RESULT 4
BJ068972          663 bp  mRNA  linear  EST 11-DEC-2001
LOCUS             BJ068972
DEFINITION        BJ068972 NIBB Mochii normalized Xenopus tailbud library Xenopus
ACCESSION         BJ068972
VERSION           BJ068972.1  GI:17497722
KEYWORDS          EST.
SOURCE            African clawed frog.
  ORGANISM        Xenopus laevis
    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
    Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;
    Xenopodinae; Xenopus.
    1 (bases 1 to 663)
    AUTHORS        Kitayama,A., Terasaka,C., Mochii,M., Ueno,N., Shin-I,T. and Kohara
    Y.
    TITLE          Expressed genes in X. laevis embryo
    JOURNAL        Unpublished (2001)
    COMMENT        Contact: Tadasu Shin-I
    National Institute of Genetics
    1111 Yata, Mishima, Shizuoka 411-8540, Japan
    Tel: 81-559-81-6856
    Fax: 81-559-81-6855
    Email: tshin@genes.nig.ac.jp.
FEATURES
  source          1..663
                  /organism="Xenopus laevis"
                  /db_xref="taxon:8355"
                  /clone_1id="XL0511f08"
                  /clone_1id="NIBB Mochii normalized Xenopus tailbud
                  library"
                  /tissue_type="whole embryo"
                  /dev_stage="stage 25"
                  /note="Vector: pBSRN3; Site_1: NotI; Site_2: EcoRI; cDNAs
                  were oligo-dT primed and directionally cloned. Staging
                  according to Nieuwkoop and Faber. Library is subtraced
                  and was constructed by N. Garrett and A.M. Zorn,
                  (Wellcome/CRC Institute)."
BASE COUNT       179 a 160 c 164 g 160 t
ORIGIN
Alignment Scores:
Pred. No.:       1.61e-09      Length: 663
Score:           181.00        Matches: 46
Percent Similarity: 42.55%     Conservative: 14
Best Local Similarity: 32.62%  Mismatches: 47
Query Match:     25.42%       Indels: 34
DB:              13           Gaps: 6
US-09-997-610-2_copy_18_149 (1-132) x BJ068972 (1-663)
QY 1 ALaGIYProProAlaHISProARgProProGIUGlUaIGlYProProGIaYAlaProGI 20
DB 106 TCAGGGCCACCGTCGTATGCAAGAGACTGCA-----GGAACACAGAGTGAAGCTGGC 156
QY 21 LeuPro-----GlnTyThrGIaUaIleSer 29
DB 157 GCTCCTGCAGAAATCGACCTACGCCCCCAGGTATGAGACAGTAATTTGCCAAATATGGGA 216
QY 30 GIuMetThrIleCysProCysProAsp-----38

```

Db	217	GCCATTGATGGAGATAAAAAACCTCATGATGATACAGGAAAAAAGAAAAAATTTGAGAGA	276
Qy	39	-----LlGluNargSerAlaPheThrValIleuSerGlyLysLeuProLeuProPhe	56
Db	277	AACATTCACGAAATGCCACCATTTCACTGCTCAACCTGACACAGCCCTTCTCTTTAGTGGGT	336
Qy	57	LysProIleIlePheThrGlyValIleuGlyrAsnIacIlnArgAspLeuLysGluAlaMet	76
Db	337	GAACTGTAAAGTTTGTAGAAACTGTTTATACACGACGAGAAATGGCTACATTCACCTTACT	396
Qy	77	GlyValPheAlaCysArgValProLysAsnGlyrTyrSerSerPheAspValGluLeuHis	96
Db	397	GGGGTGTTCACGGTAATTCACAGAAATTTATTTACTTTCTTACCATGTGC-----	447
Qy	97	HisCysLys-----ValAsnIlePheLeuMetArgLysGlnIleLeuAlaAsnLysGluGln	115
Db	448	CATTGTAAAGTGCACATGTTTGATTT-----GCGTTGTTCAGAAATATATGAGCCA	498
Qy	116	Ile Ile 116	
Db	499	GTA 501	

LOCUS	1092 bp	DNA	linear	GSS 26-JUL-2000
DEFINITION	Tetradon nigroviridis genome survey sequence T3 end of clone			
	011D22 of library A from Tetradon nigroviridis, genomic survey			
SEQUENCE				

ACCESSION	AL320996
VERSION	AL320996.1
KEYWORDS	GSS; genome survey sequence.
SOURCE	<i>Tetraodon nigroviridis</i> .
ORGANISM	<i>Tetraodon nigroviridis</i>

REFERENCE
AUTHORS

Roest Crollius,H., Jalllon,O., Dasilva,C., Bouneau,L., Fisher,C.,
1 (bases 1 to 1092)
Tetraodontidae; Tetraodon.
Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neocelosteii
1 (bases 1 to 1092)

TITLE	Estimate of human gene number provided by genome-wide analysis using Tetraodon nigroviridis DNA sequence
JOURNAL	Nat. Genet. 25 (2), 235-238 (2000)
ISSN	0969-5122

REFERENCE	2 (bases 1 to 1092)
AUTHORS	Crollius,H.R., Jallion,O., Dasilva,C., Ozouf-Costaz,C., Fizames,C.,

TITLE	JOURNAL
Characterization and repeat analysis of the compact genome of the freshwater pufferfish <i>Tetraodon nigroviridis</i>	Genome Res. 10 (7), 939-949 (2000)

REFERENCE 3 (bases 1 to 1092)

Submitted (12-APR-2000)
This sequence is a single read and was generated as part of a large

genome. For more information, please take a look at <http://www.genoscope.cns.fr/Tetraodon>.

FEATURES	Location/Qualifiers
SOURCE	1. .1092

```
/organism="Tetraodon nigroviridis"  
/db_xref="taxon:99883"
```

BASE COUNT	/note="Genoscope sequence ID : C0AA011DB1A2-end : T3"
224 a	301 c 308 g 245 t 14 others

Alignment Scores:	
Pred. No.:	1,59e-07
Score:	165.50
Percent Similarity:	45.80%
Best Local Similarity:	32.06%
Query Match:	23.24%
DB:	17
	4
length:	1092
Matches:	42
Conservative:	16
Mismatches:	32
Indels:	19
Gaps:	4

US-09-997-610-2_COPY_18_149 (1-132) X CNS0540B (1-1092)

QY 2 GLYPROGRLAHISPROARGPROPROGLUGLUGLVALGLYPROPROGLYALA----- 18

Db 528 GGGCCCCCTGGCCAGCCTGGTGCTGATGGTGACCTGGTGCCTCTGGTCT 587

19 ProglyLeuProglyInTyrThrGlyGluIle-----Ser 29

Db 588 CCCGGACCGCTGCTCCTCTGGCGAAGTTGTCTTTGAGAGAGGGAATGGGTGTGGGTGAG 647

30 GlnMetThrLysCysProCysProAspIleGluArgSerAlaPheThrValLysLeuSer 49

DB 048 G1A1GGTCAAA1CCCC-----ATG1CTGCTTCAAC1GCC1CTCTGACC 092

xy 00 01 02 03 04 05 06 07 08 09 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99

2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100 101 102 103 104 105 106 107 108 109 110 111 112 113 114 115 116 117 118 119 120 121 122 123 124 125 126 127 128 129 130 131 132 133 134 135 136 137 138 139 140 141 142 143 144 145 146 147 148 149 150 151 152 153 154 155 156 157 158 159 160 161 162 163 164 165 166 167 168 169 170 171 172 173 174 175 176 177 178 179 180 181 182 183 184 185 186 187 188 189 190 191 192 193 194 195 196 197 198 199 200 201 202 203 204 205 206 207 208 209 210 211 212 213 214 215 216 217 218 219 220 221 222 223 224 225 226 227 228 229 230 231 232 233 234 235 236 237 238 239 240 241 242 243 244 245 246 247 248 249 250 251 252 253 254 255 256 257 258 259 260 261 262 263 264 265 266 267 268 269 270 271 272 273 274 275 276 277 278 279 280 281 282 283 284 285 286 287 288 289 290 291 292 293 294 295 296 297 298 299 300 301 302 303 304 305 306 307 308 309 310 311 312 313 314 315 316 317 318 319 320 321 322 323 324 325 326 327 328 329 330 331 332 333 334 335 336 337 338 339 340 341 342 343 344 345 346 347 348 349 350 351 352 353 354 355 356 357 358 359 360 361 362 363 364 365 366 367 368 369 370 371 372 373 374 375 376 377 378 379 380 381 382 383 384 385 386 387 388 389 390 391 392 393 394 395 396 397 398 399 400 401 402 403 404 405 406 407 408 409 410 411 412 413 414 415 416 417 418 419 420 421 422 423 424 425 426 427 428 429 430 431 432 433 434 435 436 437 438 439 440 441 442 443 444 445 446 447 448 449 450 451 452 453 454 455 456 457 458 459 460 461 462 463 464 465 466 467 468 469 470 471 472 473 474 475 476 477 478 479 480 481 482 483 484 485 486 487 488 489 490 491 492 493 494 495 496 497 498 499 500 501 502 503 504 505 506 507 508 509 510 511 512 513 514 515 516 517 518 519 520 521 522 523 524 525 526 527 528 529 530 531 532 533 534 535 536 537 538 539 540 541 542 543 544 545 546 547 548 549 550 551 552 553 554 555 556 557 558 559 560 561 562 563 564 565 566 567 568 569 570 571 572 573 574 575 576 577 578 579 580 581 582 583 584 585 586 587 588 589 590 591 592 593 594 595 596 597 598 599 600 601 602 603 604 605 606 607 608 609 610 611 612 613 614 615 616 617 618 619 620 621 622 623 624 625 626 627 628 629 630 631 632 633 634 635 636 637 638 639 640 641 642 643 644 645 646 647 648 649 650 651 652 653 654 655 656 657 658 659 660 661 662 663 664 665 666 667 668 669 670 671 672 673 674 675 676 677 678 679 680 681 682 683 684 685 686 687 688 689 690 691 692 693 694 695 696 697 698 699 700 701 702 703 704 705 706 707 708 709 710 711 712 713 714 715 716 717 718 719 720 721 722 723 724 725 726 727 728 729 730 731 732 733 734 735 736 737 738 739 740 741 742 743 744 745 746 747 748 749 750 751 752 753 754 755 756 757 758 759 760 761 762 763 764 765 766 767 768 769 770 771 772 773 774 775 776 777 778 779 780 781 782 783 784 785 786 787 788 789 790 791 792 793 794 795 796 797 798 799 800 801 802 803 804 805 806 807 808 809 810 811 812 813 814 815 816 817 818 819 820 821 822 823 824 825 826 827 828 829 830 831 832 833 834 835 836 837 838 839 840 841 842 843 844 845 846 847 848 849 850 851 852 853 854 855 856 857 858 859 860 861 862 863 864 865 866 867 868 869 870 871 872 873 874 875 876 877 878 879 880 881 882 883 884 885 886 887 888 889 890 891 892 893 894 895 896 897 898 899 900 901 902 903 904 905 906 907 908 909 910 911 912 913 914 915 916 917 918 919 920 921 922 923 924 925 926 927 928 929 930 931 932 933 934 935 936 937 938 939 940 941 942 943 944 945 946 947 948 949 950 951 952 953 954 955 956 957 958 959 960 961 962 963 964 965 966 967 968 969 970 971 972 973 974 975 976 977 978 979 980 981 982 983 984 985 986 987 988 989 990 991 992 993 994 995 996 997 998 999 1000 1001 1002 1003 1004 1005 1006 1007 1008 1009 1010 1011 1012 1013 1014 1015 1016 1017 1018 1019 1020 1021 1022 1023 1024 1025 1026 1027 1028 1029 1030 1031 1032 1033 1034 1035 1036 1037 1038 1039 1040 104

[illegible][illegible]

013 075

Over 108 c[arriage]loads of w[ool] were exported in 1908

873 CCTATTATGTTCACTTATGACGAGTACCAACAC 905

RESULT 6

BJ494317	512 bp	MRNA	linear	EST	08-AUG-2002
LOCUS					

DEFINITION BC049431 / MF01FSA CDNA of *yzias latipes* CDNA clone MF01FSA024E16 5', mRNA sequence.

ACCESSION	BJ494317
VERSION	GI:22146243

SOURCE Japanese medaka.

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Acanthomorpha; Acanthopterygii; Percomorpha; Atheri
 polioformos; Atherinobranchii; Atherinomorpha; Atheri
 polioformos; Atherinobranchii; Atherinomorpha; Atheri

REFERENCE
1 (bases 1 to 512)
AUTHORS
kohara Y
shin-i T
kimura T
Nawata T
Tindo T
and Takada H

TITLE Medaka EST project in Takeda's lab
JOURNAL: unpublished (2001)

COMMENT contact: Tadasu Shin-i
Center For Genetic Resource Information

National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan

Tel: 81-339-81-0620
Fax: 81-559-81-6855

Email: csundregenes.ny.ac.jp
Location/Qualifiers

Source	1. 1912
/organism="Oryzias latipes"	

```
/db_xref="taxon:8090"
```

```
/clone_lib="MF01FSA CDNA"
```

/tissue_type="whole embryo"

RESULT 7	
LOCUS	AV593019/c
DEFINITION	AV593019 587 bp mRNA linear EST 27-NOV-2001
ACCESSION	AV593019 Bos taurus cartilage fetus Bos taurus cDNA clone
VERSION	E1CA005A05 5 , mRNA sequence.
KEYWORDS	AV593019.1 GI:9708176
SOURCE	EST.
ORGANISM	cow.
REFERENCE	Bos taurus
AUTHORS	Eukaryota; Metazoa; Chordata; Craniota; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovidae; Bovinae; Bos. 1 (bases 1 to 587) Takasuga,A., Hironosune,S., Itoh,R., Itohzono,A., Suzuki,H., Aso,H. and Sugimoto,Y.
TITLE	Establishment of a high throughput EST sequencing system using poly(A) tail-removed cDNA libraries and determination of 36,000 bovine ESTs
JOURNAL	Nucleic Acids Res. 29 (22), E108 (2001)
MEDLINE	21570554
COMMENT	Contact: Yoshikazu Sugimoto Animal Genetics Division Shirakawa Institute of Animal Genetics Oakura, Nishigo, Nishi-shirakawa, Fukushima 961-8061, Japan Tel: 81-248-25-5641 Fax: 81-248-25-5725 Email: kazusugicocoo.ocn.ne.jp Single pass sequencing. This clone was obtained from a polyA-deleted cDNA library. Location/Qualifiers 1..587
FEATURES	
source	

RESULT	B
LOCUS	BBJ18780
DEFINITION	BBJ18780 RIKEN full-length enriched, adult male corpora quadrigemina Mus musculus cDNA B230377C02 3'
ACCESSION	BBJ18780
VERSION	BBJ18780.2 GI:16403219
KEYWORDS	EST.
SOURCE	house mouse.
ORGANISM	Mus musculus
REFERENCE	Eumariota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. I (bases 1 to 689)
AUTHORS	Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A, Hiramoto,K., Horii,F., Ishii,Y., Ito,M., Kawai,J., Konno,H., Koude , M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M., Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasak , D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toya.T., Muramatsu,M. and Hayashizaki.Y.
TITLE	RIKEN Mouse ESTS (Arakawa,T., et al. 2001) Unpublished (2001)
JOURNAL COMMENT	On Jul 11, 2000 this sequence version replaced gl:9025815. Contact: Yoshihide Hayashizaki Laboratory for genome exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute The Institute of Physical and Chemical Research (RIKEN)

BASE COUNT	143 a	148 c	135 g	126 t
ORIGIN				
Alignment Scores:				
Pred. No.:	1.53e-06			552
Score:	152.50			Matches: 39
Percent Similarity:	49.58%			Conservative: 20
Best local Similarity:	32.77%			Mismatches: 53
Query Match:	21.42%			Indels: 7
DB:	10			Gaps: 2
US-09-997-610-2_COPY_18_149 (1-132) x BE625509 (1-552)				
OY 14	GLYPRGPGGLYALPProGlyLeuPProGlnInrYrThngYgLIuIleSerGIuMethrLys 33			
Db 22	GGGCCACCGGGCTTCCCGGAACCCCTGGCAGAAAGAGAGACCTGGAGAACCCGCTTAT 81			
OY 34	CysProCysProAspLIleGluArgSerAlaPheThrValLysLeuSerGIyLysLeuPro 53			
Db 82	-----GCTATACGCGTCACACGCTTCACTGTGGCGCTGGAGACCCGCTCACT 126			
OY 54	LeuProPheLysProIleIlePheThrGlyValLeuYrAsnAlaGlnArgAspLeuLys 73			
Db 127	GTTCACCAATGTACACCTCGCTTCTTAAGACTCTTCTACAAACACCAATCATTTATGAC 186			
OY 74	GLUAlaMetGIyValPheAlaCysArgValPProGIyAsnYrTYrSerSerPheAspVal 93			
Db 187	GGCAGCACTGGCAAGTCTTACTGCAACATTCGCGACTCTTACTCTTCTTACCAATC 246			
OY 94	GlueuHIS-----HIScysValAsnIleTrPLeuMetArgLysGlnIleLeuAla 111			
Db 247	ACGGGTGTCATGAAAGATGTGAAGTGTGAGCCCTTTCACAAAGACAGAGCCCTTCTTC 306			
OY 112	AsnYsgGluGluIleSerLysGlnInserIleGlnGluValYThrTrpValLeuLeu 130			
Db 307	ACCTACGACCAATATCAGGAAAGATGTGGACGACGCTGTGGCTGTGCTCTC 363			
RESULT 15				
LOCUS	BB625284			
DEFINITION	BB625284 RIKEN full-length enriched, adult male cecum Mus musculus			
ACCESSION	BB625284			
VERSION	BB625284.1			
KEYWORDS	EST.			
SOURCE	house mouse.			
ORGANISM	Mus musculus			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 633)			
AUTHORS	Arakawa,T., Carlini,C., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hiramoto,K., Horii,F., Ishii,Y., Ito,M., Kawai,J., Kono,H., Kouda,M., Koyama,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M., Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasakawa,T., Shibata,K., Shingawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toyata,T., Muramatsu,M. and Hayashizaki,Y.			
TITLE	RIKEN Mouse ESTs (Arakawa,T., et al. 2001)			
JOURNAL	Unpublished (2001)			
COMMENT	Contact: Yoshihide Hayashizaki Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute The Institute of Physical and Chemical Research (RIKEN) 1-7-22 Saito-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan Tel: 81-45-503-9222 Fax: 81-45-503-9216 Email: genome-res@sc.riken.go.jp URL: http://genome.gsc.riken.go.jp/ Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Kono,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y. Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)			

Wagii, K., Fujiiwaka, S., Inoue, K., Togawa, Y., Itawa, M., Ohara, E., Matsuki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsunaga, S., Kawai, J., Okazaki, Y., Morimatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.

RIKEN Integrated sequence analysis (RISA) system-384-format sequencing pipeline with 384 multicapillary sequencer. *Genome Res.* 10 (11), 1757-1771 (2000)

Kono, H., Fukunishi, Y., Shibata, K., Itoh, M., Carninci, P., Sugahara, Y. and Hayashizaki, Y.

Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. *Genome Res.* 11 (2), 281-289 (2001)

Kondo, S., Shingawa, A., Saito, T., Kiyosawa, H., Yamana, I., Aizawa, K., Fukuda, S., Hara, A., Itoh, M., Kawai, J., Shibata, K. and Hayashizaki, Y.

Computational Analysis of Full-length Mouse cDNAs Compared with Human Genome Sequences. *Mamm. Genome* 12, 673-677 (2001)

Please visit our web site (<http://genome.gsc.riken.go.jp>) for further details.

e mouse tissues.

Location/Qualifiers

1. .633

	/originalname="Mus musculus"
	/strain="C57BL/6J"
	/db_xref="taxon:10090"
	/clone="9130217G22"
	/clone_1bp="RIKEN full-length enriched, adult male cecum"
	/sex="male"
	/tissue_type="cecum"
	/dev_stage="adult"
	/lab_host="DH108"
	/note="Site.1: SalI; Site.2: BamHI; CDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand CDNA was primed with a primer [5':
	GAGGAGAGAGAGATCCACAGAGCTTTTTCCTTTTTTTTNN 3']
	CDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. CDNA went through one round of normalization to Rot = 10.0 and subtraction to Rot = 185.0. Second strand CDNA was prepared with the primer adapter of sequence [5' GAGGAGAGAGATTCGGATTAAATTAATTCCTCCCCCCC 3']
	CDNA was cloned into the XhoI and BamHI sites.
	vector: a modified Bluescript KS(+) after bulk excision from Lambda FLC I. Cloning sites, 5' end: SalI; 3' end: BamHI"
BASE COUNT	168 a 152 c 194 g 119 t
ORIGIN	
Alignment Scores:	
Pred. No.:	1.87e-06 Length: 633
Score:	152.50 Matches: 42
Percent Similarity:	46.77% Conservative: 16
Best Local Similarity:	33.87% Mismatches: 55
Query Match:	21.42% Indels: 11
DB:	10 Gaps: 4
US-09-997-610-2_COPY_18_149 (1-132) x BB625284 (1-633)	
OY 2	glyProProlAlHisProArgProGluValGlyPro-----15
Db 202	ggccccaatcgcccaaacagacggtccacaggagadagctgaaccattgggcccccagggag 261
OY 16	ProGlyalaProGlyLeuProGlnTyThrGlyGluIleSerGluMetThLysCys---34
	::: :: :::
Db 262	ccaggagatcagagaatagagagctggaaacgcgatcgagagagagggaaagtgggt 321
OY 35	ProCysProaspIleGluArgSerAlaPheThrValIleuSer-----GlyLeu 52
	::::: :::: :::
Db 322	gagcgtcccccttggtcccaagagtcctttactctggcacacacagctcatcagtagaacgttc 381

```
QY 53 ProLeuProPheLysProIleIlePheThrGlyValLeuTyrAsnAlaGlnArgAspLeu 72
    ||| ||| ||||| ||| :||| |||
Db 382 CFTCCCCAGATGCACCATTAATTCGATTAAGATCCTATACATGAATGAACCACTAC 441
QY 73 LysGlnAlaMetGlyValPheAlaCysArgValProGlyAsnTyrTyrSerSerPheAsp 92
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 442 AATGTAGCGCAGGGGAGATTCCACCTGCGCAGGTGCTCTATTACTTTACCTACCAT 501
QY 93 ValGluLeuHisHisCysLysValAsnIleTrpLeuMetArgLys-----GlnIleLeu 110
    :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 502 ATCACTGTGTTCTCCAGAAGATGTGCAGGTATCTTTGCTCAAAAACGGGGTAAGTCTCTG 561
QY 111 AlaAsnLysGlu 114
    |||||
Db 562 CACACCAAGAA 573
```

Search completed: February 19, 2003, 19:35:51
Job time : 1032.49 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: February 19, 2003, 14:57:18 / Search time 1839.21 seconds

(without alignments)
2088.706 Million cell updates/sec

Title: US-09-997-610-2_COPY_18_149

Perfect score: 712

Sequence: 1 AGPPAHRRPRPEVGPARGA.....KEEISKQOSIQEVTWVLLKA 132

Scoring table:

BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL=frame_p2n.model -DEV=xlh
-O=/cgn2.1/USPTO_SPOOL/US09997610/runat_10022003_160823_23819/app_query.fasta_1.1635
-DB=GenEmbl -QFMT=lastap -SUFFIX=rge -MINMATCH=0.1 -LOOPL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blomsu62 -TRANS=humand0.cdi -LIST=45
-OCALIGN=200 -THR.SCORE=PCT -THR.MAX=100 -THR.MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=eto -NORM=ext -HEAPSIZE=500 -MTLEN=0 -MAXLEN=200000000
-USER=US9997610.ecgn_1_13965_etunat_10022003_160823_23819 -NCPU=6 -ICPU=3
-NO_XLPHY -NO_MMAPP -LANG=EQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV.TIMEOUT=120
-WARN.TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

GenEmbl:*
1: gb_da:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*
15: em_da:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*

29: em_vl:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pin:*
35: em_htg_rnd:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	623	87.5	145880	9 HS302D9	282198 Human DNA s
2	320.5	45.0	1265	10 TMSHP20A	D12974 Tamias asia
3	290	40.7	5121	10 AB067779	AB067779 Tamias sl
4	217	30.5	2005	10 TMSHP25	D12975 Tamias asia
5	169.5	23.8	1918	10 RNO131848	AJ131848 Rattus no
6	167.5	23.5	2609	5 CHKX	M13496 Chicken typ
7	167	23.5	559	10 AB06781363	AB069679 Callosciu
8	167	23.5	1385	10 TMSHP27	D12976 Tamias asia
9	163.5	23.0	2235	6 AX332258	AX332258 Sequence
10	163.5	23.0	2235	9 HSCOL8A1	X57527 Human COL8A
11	163.5	23.0	2415	4 RABCOLV11I	J05042 Rabbit type
12	162.5	22.8	2506	9 BC013581	BC013581 Homo sapi
13	162.5	22.8	117000	9 AC069222	AC069222 Homo sapi
14	162.5	22.8	145880	9 HS302D9	282198 Human DNA s
15	162.5	22.8	152354	2 AC120598	AC120598 Homo sapi
16	162.5	22.8	169922	2 AC022883	AC022883 Homo sapi
17	162.5	22.8	182978	30 AC067824	AC067824 Homo sapi
18	161	22.6	7089	4 AF222861	AF222861 Sus scrof
19	159.5	22.4	1986	4 MMCOL8A	X66927 M.musculus
20	159	22.3	3143	4 BTCOL10A1	X53556 Bovine COL1
21	158	22.3	3422	10 MMA1TXCOL	221610 M.musculus
22	158	22.2	9331	10 MMCOL10A	X67348 Mus musculu
23	158	22.2	68834	2 AC119229	AC119229 Mus muscu
24	158	22.2	179303	2 AC016087	AC016087 Homo sapi
25	158	22.2	206213	2 AC021709	AC021709 Mus muscu
26	157.5	22.1	2435	10 BC011061	BC011061 Mus muscu
27	156	21.9	731	10 MMA1XCOL	X63013 M.musculus
28	156	21.9	1973	9 HS10A1COL	X65120 H.sapiens C
29	156	21.9	3215	9 HSCOLX3	X72580 Homo sapien
30	156	21.9	3226	6 AX333243	AX333243 Sequence
31	156	21.9	3226	9 HSCOLAIIX	X60382 H.sapiens C
32	156	21.9	3226	11 G28608	G28608 human STS S
33	156	21.9	10058	9 HSCOLIX	X98568 H.sapiens t
34	156	21.9	107553	9 HS0136014	AL121963 Human DNA
35	156	21.9	205594	2 AL355373	AL355373 Homo sapi
36	155.5	21.8	582	4 AFA17206	AFA17206 Canis fam
37	152.5	21.4	947	10 BC028770	BC028770 Bos tauru
38	152.5	21.4	1134	4 AF269230	AF269230 Bos tauru
39	152.5	21.4	200942	2 AC125396	AC125396 Mus muscu
40	150.5	21.1	1152	6 AX195207	AX195207 Sequence
41	150.5	21.1	1152	6 AX358517	AX358517 Sequence
42	150.5	21.1	1152	6 MMU49915	U49915 Mus musculu
43	150.5	21.1	1276	6 AR034252	AR034252 Sequence
44	150.5	21.1	1276	6 AX195209	AX195209 Sequence
45	150.5	21.1	1276	6 AX358519	AX358519 Sequence

ALIGNMENTS

RESULT 1

```
HS302D9      145880 bp      DNA      linear      PRI 12-DEC-1999
LOCUS
DEFINITION   Human DNA sequence from clone RP1-302D9 on chromosome 22 Contains
ACCESSION   GSSs, complete sequence.
VERSION      282198
KEYWORDS     282198.2 GI:6572207
SOURCE       HTG.
ORGANISM     Homo sapiens.
              Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
              1 (bases 1 to 145880)
REFERENCE    Bridgeman,A.
AUTHORS      Direct Submission
              Submitted (08-DEC-1999) Sanger Centre, Hinxton, Cambridgeshire,
              CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
              requests: clonerequest@sanger.ac.uk
COMMENT      On Dec 13, 1999 this sequence version replaced gi:3164067.
              During sequence assembly data is compared from overlapping clones.
              Where differences are found these are annotated as variations
              together with a note of the overlapping clone name. Note that the
              variation annotation may not be found in the sequence submission
              corresponding to the overlapping clone, as we submit sequences with
              only a small overlap as described above.
              This sequence has been finished according to sequence map criteria
              as follows. An attempt is made to resolve all sequencing problems,
              such as compressions and repeats, but not necessarily within known
              annotated human repeat sequence elements (e.g. Alu). Where the
              sequence is ambiguous, there is an annotation using the 'unsure'
              feature key.
              The following abbreviations are used to associate primary accession
              numbers given in the feature table with their source databases:
              Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
              on the WORMPEP database can be found at
              http://www.sanger.ac.uk/Projects/C.elegans/wormpep This sequence
              was generated from part of bacterial clone contigs of human
              chromosome 22, constructed by the Sanger Centre Chromosome 22
              Mapping Group. Further information can be found at
              http://www.sanger.ac.uk/HGP/Chr22
              RP1-302D9 is from the library RPCI-1 constructed at the Roswell
              Park Cancer Institute by the group of Pieter de Jong. For further
              details see http://bacpac.med.buffalo.edu/
              VECTOR: pcypac2
              This sequence is the entire insert of clone RP1-302D9 The true left
              end of clone CTA-282F2 is at 69682 in this sequence. The true right
              end of clone CTA-41562 is at 55167 in this sequence.
FEATURES
source       1..145880
              /organism="Homo sapiens"
              /db_xref="taxon:9606"
              /chromosome="22"
              /clone="RP1-302D9"
              /clone_lib="RPCI-1"
              188..245
              /note="MIR3 repeat: matches 144..209 of consensus"
              repeat_region
              246..571
              /note="AluX repeat: matches 1..312 of consensus"
              repeat_region
              572..759
              /note="MIR3 repeat: matches 1..144 of consensus"
              repeat_region
              763..933
              /note="MIR5A repeat: matches 26..187 of consensus"
              repeat_region
              1033..1336
              /note="AluSp repeat: matches 1..299 of consensus"
              repeat_region
              1450..1583
              /note="MIR repeat: matches 24..160 of consensus"
              repeat_region
              1687..1752
              /note="L2 repeat: matches 2593..2661 of consensus"
              repeat_region
              2350..2660
              /note="AluSc repeat: matches 3..309 of consensus"
              repeat_region
              2684..2981
              /note="AluSq repeat: matches 2..300 of consensus"
              repeat_region
              3323..3343
              /note="MIR repeat: matches 116..136 of consensus"
              repeat_region
              3344..3652
              /note="AluY repeat: matches 1..309 of consensus"
              repeat_region
              3653..3928
              /note="MIR1E repeat: matches 136..359 of consensus"
              repeat_region
              3929..4278
              /note="MIR1B repeat: matches 3..364 of consensus"
              repeat_region
              4279..4485
              /note="MIR1E repeat: matches 359..568 of consensus"
              repeat_region
              5073..5176
              /note="52 copies 2 mer ct 78 conserved"
              repeat_region
              5181..5491
              /note="AluJb repeat: matches 1..311 of consensus"
              repeat_region
              6369..6485
              /note="L2 repeat: matches 2579..2705 of consensus"
              repeat_region
              6647..6685
              /note="MIR1 repeat: matches 1..23 of consensus"
              repeat_region
              6686..6987
              /note="AluX repeat: matches 1..302 of consensus"
              repeat_region
              6988..7036
              /note="MIR1 repeat: matches 23..77 of consensus"
              repeat_region
              7482..7754
              /note="AluJb repeat: matches 9..290 of consensus"
              repeat_region
              7775..8060
              /note="AluJo repeat: matches 1..295 of consensus"
              repeat_region
              8414..8551
              /note="L2 repeat: matches 2553..2706 of consensus"
              repeat_region
              8914..9030
              /note="MIR repeat: matches 147..262 of consensus"
              repeat_region
              9110..9280
              /note="MIR repeat: matches 91..262 of consensus"
              repeat_region
              9283..9412
              /note="MIR repeat: matches 15..144 of consensus"
              repeat_region
              9521..9679
              /note="FAM repeat: matches 3..161 of consensus"
              repeat_region
              9820..10225
              /note="MIR repeat: matches 2..425 of consensus"
              repeat_region
              complement(10179..10678)
              /note="match: GSS: Em:B56592"
              misc_feature
              complement(10204..10728)
              /note="match: GSS: Em:AQ071486"
              misc_feature
              complement(10249..10706)
              /note="match: GSS: Em:AQ25495"
              repeat_region
              10312..10383
              /note="MIR repeat: matches 79..150 of consensus"
              repeat_region
              10718..11310
              /note="match: GSS: Em:BL4024"
              misc_feature
              10784..11201
              /note="match: GSS: Em:B43656"
              repeat_region
              11838..11946
              /note="MIR repeat: matches 20..137 of consensus"
              repeat_region
              12174..12445
              /note="L2 repeat: matches 1988..2275 of consensus"
              repeat_region
              12444..12642
              /note="MIR repeat: matches 63..241 of consensus"
              repeat_region
              13017..13369
              /note="match: STS: Em:G49301"
              misc_feature
              13331..13397
              /note="MIR repeat: matches 174..244 of consensus"
              repeat_region
              13398..13698
              /note="MIR repeat: matches 1..302 of consensus"
              repeat_region
              13699..13810
              /note="MIR repeat: matches 76..174 of consensus"
              repeat_region
              13806..13919
              /note="MIR repeat: matches 77..189 of consensus"
              repeat_region
              13945..14060
              /note="MIR repeat: matches 24..142 of consensus"
              repeat_region
              14061..14367
              /note="AluY repeat: matches 1..301 of consensus"
              repeat_region
              14368..14452
              /note="MIR repeat: matches 141..225 of consensus"
              repeat_region
              14589..14679
              /note="MIR repeat: matches 173..262 of consensus"
              repeat_region
              14597..15201
```

```

misc_feature      /note="match: GSS: Em:AQ553482"
14616..15060
/note="match: GSS: Em:AQ370601"
repeat_region     14868..15040
/note="MIR repeat: matches 49..233 of consensus"
repeat_region     15071..15188
/note="L2 repeat: matches 2112..2239 of consensus"
repeat_region     15304..15399
/note="MT1B repeat: matches 1..99 of consensus"
repeat_region     15490..15662
/note="AluSgl repeat: matches 2..114 of consensus"
repeat_region     15669..15727
/note="MT1B repeat: matches 119..178 of consensus"
repeat_region     15728..16027
/note="AluSc repeat: matches 1..299 of consensus"
repeat_region     16028..16245
/note="MT1B repeat: matches 178..390 of consensus"
repeat_region     16346..16854
/note="AluY repeat: matches 1..300 of consensus"
repeat_region     18296..18323
/note="MSTA repeat: matches 2..29 of consensus"
repeat_region     18324..18392
/note="MER66-internal repeat: matches 4919..4993 of consensus"
repeat_region     18393..18712
/note="AluYb repeat: matches 1..311 of consensus"
repeat_region     18713..19133
/note="MER66-internal repeat: matches 4548..4919 of consensus"
/note="complement(18872..19230)"
misc_feature      /note="match: GSS: Em:AQ005063"
19251..19719
/note="match: GSS: Em:B14179"
repeat_region     19537..20280
/note="HERVFN21 repeat: matches 4657..5784 of consensus"
repeat_region     20317..20382
/note="33 copies 2 mer ta 68 conserved"
repeat_region     20513..20666
/note="77 copies 2 mer tt 70 conserved"
repeat_region     20682..21008
/note="AluSg1 repeat: matches 1..306 of consensus"
repeat_region     21239..21553
/note="HURS-P3 repeat: matches 4410..4713 of consensus"
repeat_region     21882..22254
/note="THB1B repeat: matches 1..364 of consensus"
repeat_region     22302..22537
/note="MER66-internal repeat: matches 2186..2417 of consensus"
repeat_region     22538..22850
/note="AluSp repeat: matches 1..313 of consensus"
repeat_region     22851..23801
/note="MER66-internal repeat: matches 1210..2186 of consensus"
repeat_region     23905..23989
/note="MER66-internal repeat: matches 3017..3102 of consensus"

Alignment Scores:
Pred. No.: 6.49e-47 Length: 145880
Score: 623.00 Matches: 117
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 87.50% Indels: 0
DB: 9 Gaps: 0

US-09-997-610-2_COPY_18_149 (1-132) x HS302D9 (1-145880)
QY 16 ProGlyAlaProGlyLeuProGlyIleThrGlyGluIleSerGlyMetThrIlyCysPro 35
DB 37262 CCTGGTGGCCCGAGTTTACCAACAATATACAGGAATATAGTGAATGACAAATGCCCC 37321
QY 36 CysProAspIleGluArgSerAlaPheThrValIlySleSerGlyIlyLeuProLeuPro 55
DB 37322 TGGCGGTATATGAAAGGTACAGCCTTTACTGTGAAGCTCAGTGGAAACTCTCTTCC 37381

```

```

QY 56 PheIysProIleIlePheThrGlyValLeuTyraSAlaInArGAspleuIysGluAla 75
DB 37382 TTCAGCCCATCATCTTACACAGGGGTCCTGTACAAATGCCAGAGGGATTAAAGAGGCC 37441
QY 76 MetGlyValPheAlaCysArgValProGlyAsnTyTyrSerSerPheAspValGluLeu 95
DB 37442 ATGGAGTCTTTCCTTGCAGGGTCCCTGGGAATTAATCTCCAGCTTTGATTTGAGCTG 37501
QY 96 HisHisCysIysValAsnIleTrpLeuMetArgIlySgIleLeuAlaAsnIysGluGlu 115
DB 37502 CATCATTTGCAAGGTGATATTTGGCTPATGAGGAACAATTTGGCTAATAAGAGAA 37561
QY 116 IleSerIysGlnIleSerIleGlnIleGluValThrTrpValLeuLeuIysAla 132
DB 37562 ATTTCATACGAGCAAGCATTAAGAGGTGACTGGTGCTGTATAAGCA 37612
RESULT 2
TMSHP20A 1265 bp mRNA linear ROD 03-FEB-1999
LOCUS TMSHP20A
DEFINITION Tamias asiaticus mRNA for HP-20, complete cds.
ACCESSION D12974
VERSION D12974.1 GI:287467
KEYWORDS HP-20; collagen-like domain; hibernation-related protein; plasma protein.
SOURCE Tamias asiaticus liver cDNA to mRNA, clone:PCM20-7.
ORGANISM Tamias sibiricus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sclurognathi; Scluridae; Sclurinae;
Tamias.
REFERENCE 1 (bases 1 to 1265)
AUTHORS Takamatsu N., Ohba K., Kondo J., Kondo N. and Shiba T.
TITLE Hibernation-associated gene regulation of plasma proteins with a
JOURNAL Mol. Cell. Biol. 13 (3), 1516-1521 (1993)
MEDLINE 93180798
REFERENCE 2 (bases 1 to 1265)
AUTHORS Takamatsu N.
TITLE Direct Submission
JOURNAL Submitted (21-AUG-1992) Nobuhiko Takamatsu, School of Hygienic
Sciences, Kitasato University, Laboratory of Molecular Biology;
1-15-1 Kitasato, Sagamihara, Kanagawa 228, Japan (Tel:0427-78-9408,
Fax:0427-78-9403)
FEATURES
source location/Qualifiers
1..1265
/organism="Tamias sibiricus"
/db_xref="taxon:64680"
/clone="PCM20-7"
/tissue_type="liver"
89..679
/note="collagen-like domain at nt 161-277"
/codon_start=1
/product="HP-20"
/protein_id="BAA02351.1"
/db_xref="GI:287468"
/translacion="MTDVRLALITFLMVLVNDQVSCSPRPVGYPGVYGRGP
PQPGAGRPGDGPSPKPCRSRSLFTVKFGRLPRPSPVAVFVLYNTQRL
KSTGVNVEPQGNHFSFDELVLHCKVIGLMQVEMKHOLSKNEYENASGAMI
MPLRGDKVWLADVETEPDQAKVIVYSGFLISS"
89..157
sig_peptide
mat_peptide
158..676
/product="HP-20"
1238..1243
polyA_signal
polyA_site
1265
BASE COUNT 354 a 304 c 288 g 319 t
ORIGIN
Alignment Scores:
Pred. No.: 6.29e-21 Length: 1265
Score: 320.50 Matches: 66
Percent Similarity: 65.00% Conservative: 12
Best Local Similarity: 55.00% Mismatches: 39
Query Match: 45.01% Indels: 3

```

```
DB: 10 Gaps: 2
US-09-997-610-2_Copy_18_149 (1-132) x TMSHP20A (1-1265)
QY 2 GlyProProAlaHisProArgProProGluGluValAlGlyProProGlyAlaProGlyLeu 21
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 188 GGAGTGCCTCCGGAGTCCGGGGCTCTGGGGCCCCCAGAGTACACAGGCGTGGCGGTAGG 247
QY 22 ProGlnIYrThrGlyGluIleSerGluMetThrLysCysProCysProAspIleGluArg 41
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 248 CCAGAGACCCAGACCCAAAAGGCCACAGTGAATAATGCCCTGCAGCA-----GAGAGG 301
QY 42 SerAlaPheThrValLysLeuSerGlyLysLeuProLeuProPheLysProIleIlePhe 61
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 302 TCAGCCTTCACGGTCAAGTTCACGGAGAGGCTCCCTCCACCTCGGAGCCTGTGGCTTTC 361
QY 62 ThrGlyValLeuThrValAsnAlaGlnArgAspLeuLysGluAlaMetGlyValAlaPheAlaCys 81
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 362 ACAGAGCTCTGTACATATACCCAGAGGACTTGAAGAGACAGACAGAGCTTTAACTGC 421
QY 82 ArgValProGlyAsnIYrThrSerPheAspValGluLeuHisCysLysValAsn 101
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 422 GTGAGCCTCGAATAATACCATTTACAGCTTGTATGTTGAGCTTACACCTGCAGAGGTGAAG 481
QY 102 IleTyrLeuMetArgLysGlnIle---LeuAlaAsnLysGluGluIleSerLysGlnGln 120
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 482 ATTTGGGTTGTATGAAGAACAACATCAAGTCAATGAAAAAGCATCAGCTTCCAAAATGAA 541

RESULT 3
AB067779 5121 bp DNA linear ROD 26-OCT-2001
LOCUS Tamias sibiricus gene for Hp-20, complete cds.
DEFINITION AB067779
ACCESSION AB067779
VERSION AB067779.1 GI:15706341
KEYWORDS Tamias sibiricus
SOURCE Tamias sibiricus DNA, clone:lamdacM2061.
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Sciuridae; Sciurinae;
Tamias.
REFERENCE 1
AUTHORS Ono,M., Hosoe,Y., Azuma,S., Shoji,M., Nara,K., Kondo,N., Shiba,T.
and Takamatsu,N.
TITLE HNF-1 regulates the liver-specific transcription of the chipmunk
HP-20 gene
JOURNAL Gene 277 (1-2), 121-127 (2001)
MEDLINE 21488336
REFERENCE 2 (bases 1 to 5121)
AUTHORS Ono,M.
TITLE Direct Submision
JOURNAL Submitted (04-AUG-2001) Motoharu Ono, Kitasato University,
Department of Biosciences, School of Science, 1-15-1 Kitasato,
Sagamihara, Kanagawa 228-8555, Japan
(E-mail:ms99805@estu.sci.kitasato-u.ac.jp, Tel:81427789408,
Fax:814277789408)

FEATURES
source Location/Qualifiers
1..5121
/organism="Tamias sibiricus"
/db_xref="taxon:64680"
/clone="lamdacM2061"
/note="synonym: Tamias asiaticus"
1518..1684
/join(1678..1684,2573..2710,3968..4413)
/gene="HP-20"
/join(1678..1684,2573..2710,3968..4413)
/gene="HP-20"
/codon_start=1
/product="HP-20"
/protein_id="BAB68362.1"
/db_xref="GI:15706342"
/translation="MTDAARLAFVLMNVNLDVYSCSGPPGVGPGVYGRGP
PGQPCAGRPDGPGRKPSVCKPCRSRAFTVKGSRLEPPEPVFEVLVNTQDRL
KASTGVNCEPGENYHFSFDVELYHCKVKGILMKNHIOVMEKHOLSKNEYENASGAMI
```

```
exon MFLRGDKYWLADVETEEPDQAKVVIYFSGFLISS"
2573..2710
/gene="HP-20"
exon 3968..5000
BASE COUNT 1571 a 1018 c 1090 g 1442 t
ORIGIN
Alignment Scores:
Pred. No.: 2,01e-17 Length: 5121
Score: 290.00 Matches: 60
Percent Similarity: 66.97% Conservative: 13
Best Local Similarity: 55.05% Mismatches: 30
Query Match: 40.73% Indels: 6
DB: 10 Gaps: 3
US-09-997-610-2_Copy_18_149 (1-132) x AB067779 (1-5121)
QY 13 ValGlyProProGlyAlaProGlyLeuProGlnIYrThrGlyGluIleSerGluMetThr 32
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3964 CTAGGCGCTCGGGGTAGGACGAGACCA-----GGACCAAAAAGGCCCACTGTA 4014
QY 33 LysCysProCysProAspIleGluArgSerAlaPheThrValLysLeuSerGlyLysLeu 52
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4015 AATGCCCCCTGCACA-----GAGAGTCAAGCTTCAAGGTGAAGTTCACAGCGAGAGCTC 4068
QY 53 ProLeuProPheLysProIleIlePheThrGlyValLeuThrValAsnAlaGlnArgAspLeu 72
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4069 CTTCCACCTTCGGAGCCTGTGGCTTTCACAGAGCTCTGTACAAATACCCAGAGGAGCTTG 4128
QY 73 LysGluAlaMetGlyValAlaPheAlaCysArgValProGlyAsnIYrThrSerPheAsp 92
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4129 AAGCGGACACAGAGAGCTTTAACTCGGTGAGCGAGCTGGAAATTAACATTCAGCTTTGAT 4188
QY 93 ValGluLeuHisCysLysValAsnIleTyrLeuMetArgLysGlnIle---LeuAla 111
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4189 GTTGAGCTTTACCACTCAAGGTGAAGATTGGTTGATGAAGAACAACATCAAGTCAATG 4248
QY 112 AsnLysGluGluIleSerLysGlnGln 120
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4249 GAAAGCATCAGCTCTCCAAAACGAA 4275

RESULT 4
TMSHP25
LOCUS TMSHP25 2005 bp mRNA linear ROD 03-FEB-1999
DEFINITION Tamias asiaticus mRNA for Hp-25, complete cds.
ACCESSION D12975
VERSION D12975.1 GI:287469
KEYWORDS HP-25; collagen-like domain; hibernation-related protein; plasma
protein.
SOURCE Tamias asiaticus liver cDNA to mRNA, clone:PCM25-3.
ORGANISM Tamias sibiricus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Sciuridae; Sciurinae;
Tamias.
REFERENCE 1 (bases 1 to 2005)
AUTHORS Takamatsu,N., Ohba,K., Kondo,J., Kondo,N. and Shiba,T.
TITLE Hibernation-associated gene regulation of plasma proteins with a
collagen-like domain in mammalian hibernators
JOURNAL Mol. Cell. Biol. 13 (3), 1516-1521 (1993)
MEDLINE 93180798
REFERENCE 2 (bases 1 to 2005)
AUTHORS Takamatsu,N.
TITLE Direct Submision
JOURNAL Submitted (21-AUG-1992) Nobuhiko Takamatsu, School of Hygienic
Sciences, Kitasato University, Laboratory of Molecular Biology,
1-15-1 Kitasato, Sagamihara, Kanagawa 228, Japan (Tel:0427-78-9408,
Fax:0427-78-9403)

FEATURES
source Location/Qualifiers
1..2005
/organism="Tamias sibiricus"
/db_xref="taxon:64680"
/clone="PCM25-3"
/issue_type="liver"
```


D6 661 AAGGGGACTCACGTTGGGTAGCGCTGTATTAAGATTGCCA-----CGCCACACA 708

OY 115 uIleSerLySGInGIInSerLIeGLInGLuValThrTrpValLeuLys 131
: : : : : ||| : : | ||||| : : | ||| : :
D6 709 ATGTACACGTATGATGAGTACAGCAAAAGCTACTTGATCAGGCTTCAG 757

RESULT 6

LOCUS	CHKCX	2609 bp	DNA	linear	VRT 28-APR-1993
DEFINITION	Chicken type X collagen gene.				
VERSION	M13496				
KEYWORDS	G1:211699				
SOURCE	Chicken red blood cell DNA, clone pYN92EL, and embryo chondrocyte, cDNA to mRNA, clone pYN3116.				
ORGANISM	Gallus gallus				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.				
AUTHORS	Nimwya, Y., Gordon, M., van der Rest, M., Schmid, T., Linsemanier, T. and Olsen, B.R.				
TITLE	The developmentally regulated type X collagen gene contains a long open reading frame without introns				
JOURNAL	J. Biol. Chem. 261 (11), 5041-5050 (1986)				
PUBMED	86168227				
COMMENT	(1) has as yet not determined the start codon of the collagen X gene. The open reading frame starts at position 166. There are several 'atg' codons (starting at positions 190, 214, 256, 274 and 310) following the 'taa' at position 163. Location/Qualifiers 1..2609 /organism="Gallus gallus" /db_xref="taxon:9031" /cell_type="chondrocyte" <1..379 <380..2208 /codon_start=3 /product="type X collagen" /protein_id="AAA48736.1" /db_xref="GI:211700" /translation="KPGFGSGPQGPPEPLCPDPGFSTVYVKIGMPGLPKKGEEKLNGE KGEAVPLGPEAKRPQGPPGIPGAGISLVLEKPPDQGPAGQGRGPRGEGEVPEP IINGCKMGEGFVGPRPNRNRLPGPDGQGLPGSAGIGKPGNLPGCPGMKGDGLPQ ARGAGILGEPDGPPEGVEVIGIKRPMGPPGPAIPGAKLIPGAPLGSPGLRGFGES RKPJLGMKGNHRRGEPGPPGPPGPKDDQAGVPGLGAPGPGNNGPGLKGLPEENG I PGPGMDGPPVCRPAFPKAGKRGLGLDGKRGYCEGDLPERKHRLPGOKGPTGAR GHEDLPKPVPOGVKGYVGLINGEVRKPSGIVKRGITGPPKMAGAGAAGEAFPE LPGAAGIVTKGLRPMGLPGRPGRKNSGEGLDPPDPGPSQTIPGVYGESER ELSSGFMKAGANQALGMPASFTVLISKAYPGATVP IKFDK ILYNQOHYDPRTHI FTTCIRGLYYFSYVSHAKGTVMWVALKNKSVPVTYDEYKGYLDOPQAVLSLSMR TIKGSSSQQLDNFVNPIKNFILLOSVELSKSNPIPLMS"				
variation	/note="a in DNA; t in cDNA"				
BASE COUNT	729 a 635 c 659 g 586 t				
ORIGIN	1009 bp upstream of BamHI site.				
Alignment Scores:					
Pred. No.:	2,59e-06	Length:	2609		
Score:	167.50	Matches:	49		
Percent Similarity:	42.41%	Conservative:	18		
Best Local Similarity:	31.01%	Mismatches:	56		
Query Match:	23.53%	Indels:	35		
DB:	5	Gaps:	7		

Db	1699	ATCCAGAGAGCGTTATGTTAAAGGAGCGCTGAGACACTATAGCAATGCTCTTTCATGAAA	1758
OY	31	-----methrhyScyProcysProaSpIleGluArgSerAlaPhe	44
Db	1759	GCAGAGCAAAATCAAGCGCTCTACAGAGATGCCAGTG-----TCGCGCTTC	1803
OY	45	ThrValLysLeuSerGlyLysLeuProLeuProPheLysProIleIlePheThrGlyAla	64
Db	1804	ACTGCACTATCTCTCAAAAGCCCTACCCCTGGGGCAACAGTCACCAATCAAAATTTGACAAAATC	1863
OY	65	LeuTrpAsnAlaGlnArgAspLeuLysGluAlaMetGlyAlaPheAlaCysArgValPro	84
Db	1864	TTGTACAAACAGACACCAACACTATGACCCCAAGACAGCAAACTTTTACTCTGCAGATCCCT	1923
OY	85	GlyAsnTrpTrpSerSerPheAspAlaGluLeuHisCysLys--ValAsnIleTrp	103
Db	1924	GGCTATATCTATTTCTCCATCATGTGA-----CATGCAAAAAGAAACAATGTTTGG	1974
OY	104	LeuMet-----ArgLysGlnIleLeuAlaAsnLysGluGluIleSerLys---	118
Db	1975	GTTGCACCTCTATMAAATAGCGCTCCCACTGCTGATCTTATGATGATACCAAAAAGA	2034
OY	119	-----GlnGlnSerIleGlnGluValIThrTrpAlaLeuLys	131
Db	2035	TACCTTGACCAAGCCCTCAGGCACTGCTGTCATTTGATCTATGAGACAGATCAAG	2088
RESULT 7			
AB06781363			
LOCUS	AB06781353	559 bp	DNA
DEFINITION	Callosciurus caniceps HP-25 pseudogene, exon 3, complete cds.		Linear ROD 18-JAN-2002
ACCESSION	AB069679		
VERSION	AB069679.1	GI:18149913	
KEYWORDS			
SEGMENT			
ORGANISM			
REFERENCE			
AUTHORS	1 Kojima, M., Shiba, T., Kondo, N. and Takamatsu, N.		
TITLE	The tree squirrel HP-25 gene is a pseudogene		
JOURNAL	Eur. J. Biochem. 268 (22), 5997-6002 (2001)		
MEDLINE	21579740		
REFERENCE	2 (bases 1 to 559)		
AUTHORS	Takamatsu, N.		
TITLE	Direct Submission		
JOURNAL	Submitted (06-AUG-2001) Nobuhiko Takamatsu, Kitasato University,		
	Department of Biosciences, School of Science; 1-15-1 Kitasato,		
	Sagamihara, Kanagawa 228-8555, Japan		
	(E-mail: itakamatu@jet.sci.kitasato-u.ac.jp, Tel: 81427789408,		
	Fax: 81427789408)		
FEATURES			
source	Location/Qualifiers		
	1..559		
	/organism="Callosciurus caniceps"		
	/db_xref="taxon:64664"		
	/clone="lambda TS25G1"		
	join(AB067813.1:1503..1658,AB069678.1:26..185,39..559)		
gene	/gene="HP-25"		
	join(AB067813.1:1503..1658,AB069678.1:26..185,39..496)		
CDS	/gene="HP-25"		
	/pseudo		
	/codon_start=1		
	/product="HP-25"		
	39..>559		
exon	/gene="HP-25"		
	/number=3		
	/pseudo		
BASE COUNT	150 a 128 c 141 g 140 t		
ORIGIN			

Pred. No.:	5,43e-07	Length:	559
Score:	167.00	Matches:	42
Percent Similarity:	47.06%	Conservative:	14
Best Local Similarity:	35.29%	Mismatches:	61
Query Match:	23.46%	Indels:	2
DB:	10	Gaps:	1

US-09-997-610-2_COPY_18_149 (1-132) x AB067813S3 (1-559)

QY	14	GIYPProGlyAlaProGlyLeuProGlnIrrThrgIguILesSerGlumethrlys	33
Db	38	GGACATCCAGGTATACCAAGGCGAGCATGAGAGGCCCTCCTGGAGACATGTGAGCG	97
QY	34	CysProCySProAspIleGluArgSerAlapherThrValLysLeuSerGilyLysLeuPro	53
Db	98	TGCCCATCACCA-----CCAAATATTGCCTTTGGCCGTGAAGCTGAGTGAAGCAGTCGCCCA	151
QY	54	LeuProPhelySProlleIlePheThcIlyValLeuIryrAsnAglanArgaspLeuys	73
Db	152	GGGTCCTCCAGGCCCATATATTCCAAGGAGTCTGTGACACCAGGAGACCACTATATAC	211
QY	74	GIUAImetGlyValIphealAcysArgValAProGlyasnTyrrTySerSepheaspval	93
Db	212	GTGACCACTGGAGAGTTCACGTATACCAACCCGGTGTCACAAATTGGCTTGACATT	271
QY	94	GIUleuHISHSICySLysValasNIletRpleuMetargLysGInIleLeualaIsnlys	113
Db	272	GGACTGTTCAGAGTTCCTGTGAAGATAAGTCTCATGAGAGATGTATCCAGATCAGGGAG	331
QY	114	GIUGIUleSerLysGInGlnSerIleGInGUalVIMTrTPvalLeuLeuLysAla	132
Db	332	AAGCAAGCTCAAGCCATGTAAGCTACAACAATGCATTGGGAGCTGATTTGGCT	388

RESULT 8	LOCUS	TMSP27	1385 bp	mRNA	Linear	ROD 03-FEB-1999
DEFINITION	Temias asiaticus mRNA for Hp-27, complete cds.	TMSP27				
ACCESSION	D12976	D12976.1	GI:287471			
VERSION						
KEYWORDS	HP-27; collagen-like domain; hibernation-related protein; plasma protein.					
SOURCE	Temias asiaticus liver cDNA to mRNA, clone:pCM27-3.					
ORGANISM	Temias sibiricus					
REFERENCE	Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Scuridae; Sciurinae; Temias.					
AUTHORS	1 (bases 1 to 1385) Takamatsu,N., Ohba,K., Kondo,J., Kondo,N. and Shiba,T.					
TITLE	Hibernation-associated gene regulation of plasma proteins with a collagen-like domain in mammalian hibernators					
JOURNAL	Mol. Cell. Biol. 13 (3), 1516-1521 (1993)					
MEDLINE	93180798					
REFERENCE	2 (bases 1 to 1385) Takamatsu,N. Direct Submission Submitted (21-AUG-1992) Nobuhiko Takamatsu, School of Hygienic Sciences, Kitasato University, Laboratory of Molecular Biology, 1-15-1, Kitasato, Sagamihara, Kanagawa 228, Japan (Tel:0427-78-9408, Fax:0427-78-9403)					
AUTHORS						
TITLE						
JOURNAL						

FEATURES	source
Location/Oualifiers	
1..1385	/organism="Temias sibiricus"
	/db_xref="taxon:64680"
	/clone="pCM27-3"
	/tissue_type="liver"
65..712	/note="collagen-like domain at nt 191-307"
	/codon_start=1
	/product="Hp-27"
	/protein_id="BAA02353.1"
	/db_xref="GI:287472"
	/translation="MYEGRASPTMGAGITWILLALSYLMHYVCSSETGNBPSCNPPGP OGPPGMGPDPCTGKPGPDMWNNGPPGLGPPGPGPMVNCHSKTSFAVKANKLPAPA

sig_peptide	PSDPIYFKRALIDAGCHPRLATGVFTCPVGLYGFQFPHLEAVQRAVYVSLMRNCTQYH				
mat_peptide	EREBAQDGTETIHISGTAILQLQHEKDRVWLENKSLQDLEKGTQDAVFSGLIHEN"				
polya_signal					
polya_site	1369. .1374				
BASE COUNT	439 a 285 c 311 g 350 t				
ORIGIN					
Alignment Scores:					
Pred. No.:	1.45e-06				
Score:	167.00				
Percent Similarity:	43.758				
Best Local Similarity:	35.166				
Query Match:	23.466				
DB:	10				
US-09-997-610-2_COPY_18_149 (1-132) x TMSHP27 (1-1385)					
QY	2	GIYPRPFOALAHISPRQATRPPOGLUGLUVAlGlyProPro-----GLyAla	18		
DB	218	GGTCCCCCAGGTACACTGGAAAAACA-----GGCCACACAGATGGAATGGTTTT	268		
QY	19	PROGLYLeuPRoGLInTyThrGIyGLUIdSerGIuMeThrLysCysPRoCysPRoASP	38		
DB	269	CCAGGACTACAGGGCCACCAGGCCACACAGGCATGACTGTGAACSTGCCACAGCAAGGA	328		
QY	39	ILeGLIArgSerIaPheThrValLysLysSerGIyLysLeuPRoLeuPRoPheLysPro	58		
DB	329	-----ACATCGGCTTTGCAGTGAAGCAAAATGAGCTGCCCCAGCTCCCTCCAGGCC	382		
QY	59	ILeIlePheThrGIyValLeuTyTAsnAlAGIlaArgAspLeuLysGIuAlaMeTGIyAl	78		
DB	383	GGATGTTTCAAGAGAACCCCTGCATGACGCTCAGGAGACATTGATCTGGCCACTGGTGTG	442		
QY	79	PheAlaCysArgValPRoGLIAsnTyTyTyrSerSerPheAspValGIuLeuHIScys	98		
DB	443	TTCACCTCGCCAGCTCCAGGACTCTCCACATTTGGATTTCATCATTTGAAGCTGTCCAGAG	502		
QY	99	LysValAsnILeThrPheuMeTArgLysGLILeuleuLAsnLysGIuGLIILeSerLys	118		
DB	503	GCTGTAGAGGTAGCCTCATGAGAAATGGCACCCAACTCATGTGGAGAGAGAGCGGAGGCC	562		
QY	119	GIInSerILeGLInGLUValThr	126		
DB	563	CAGATGGCTATGACGACATTTCa	586		
RESULT 9	AX332258	2235 bp	DNA	linear	PAT 09-JAN-2002
LOCUS	AX332258				
DEFINITION	Sequence 2767 from Patent WO0194629.				
ACCESSION	AX332258				
VERSION	AX332258.1	GI:1812892			
KEYWORDS					
SOURCE					
ORGANISM	human.				
REFERENCE	Human sapiens				
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.				
TITLE	Young, P.E., Augustus, M., Carter, K.C., Ebner, R., Endress, G., Horrigan, S., Soppet, D.R. and Weaver, Z.				
JOURNAL	Cancer gene determination and therapeutic screening using signature gene sets				
FEATURES	Patent: WO 0194629-A 2767 13-DEC-2001;				
source	Avalon Pharmaceuticals (US)				
BASE COUNT	524 a 632 c 693 g 386 t				
ORIGIN					

SOURCE	CellBelt (strain New Zealand white) young adult cornea endothelial cells, cDNA to mRNA, clone pCE1230 and NK1.
ORGANISM	Oryctolagus cuniculus
REFERENCE	Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
AUTHORS	Yamauchi, N., Benay, P. D., van der Rest, M. and Ninomiya, Y.
TITLE	1 (bases 1 to 2415) The cloning and sequencing of alpha 1(VIII) collagen cDNAs demonstrate that type VIII collagen is a short chain collagen and contains triple-helical and carboxyl-terminal non-triple-helical domains similar to those of type X collagen
JOURNAL	J. Biol. Chem. 264 (27), 16022-16029 (1989)
MEDLINE	89380199
PUBMED	2476437
COMMENT	Draft entry and computer-readable copy of sequence [1] kindly submitted by Y. Ninomiya 23-AGC-89.
FEATURES	Location/Qualifiers
SOURCE	1..2415
CDS	/organism="Oryctolagus cuniculus" /db_xref="taxon:9986" 181..2415 /note="alpha-1 (VIII) collagen precursor" /codon_start=1 /protein_id="AA31204.1" /db_xref="GI:164896" /translation="MAVPPGPOLLQVLTLSIGSTIRLIDAGAYGKPLPPQIPPM PPIOTPIQGVPHMPLAKDGLTMKEMPHAOYKEHYHLDPQYMEVQVPMKKE AVRKKEIRPLASLRGEGRGERGEGRGEPGLPGQIGPLGKPGQGVGVK GAMKPGKPGAMKPGKAGEIGKGEIGPMGIRGPQRPDPHGLIGKPGPLPGQ PGKGDGKPGKPGPGLQSGKPGKPGGLGLKPKPMHGPQGVGLGVKPGPG TGKPGKGLKPGKPGKPGKPGKPGKPGKPGKPGKPGKPGKPGKPGKPGKPG EGGLPLGKPGKPGKPGKPGKPGKPGKPGKPGKPGKPGKPGKPGKPGKPGK GLPLPGKPGKPGKPGKPGKPGKPGKPGKPGKPGKPGKPGKPGKPGKPGK IRGLPLGKPGKPGKPGKPGKPGKPGKPGKPGKPGKPGKPGKPGKPGKPGK ICGPIGKPGKPGKPGKPGKPGKPGKPGKPGKPGKPGKPGKPGKPGKPGK PGKPGKPGKPGKPGKPGKPGKPGKPGKPGKPGKPGKPGKPGKPGKPGKPG PEPPVPGKPGKPGKPGKPGKPGKPGKPGKPGKPGKPGKPGKPGKPGKPG EPVMTYDDEKKGFLDQASAVALLPGRVFLQMPSEDAAGLYAGQVHSSPSGL LYPM"
BASE COUNT	536 a 719 c 768 g 392 t
ORIGIN	
sig_peptide	181..240 /note="alpha-1 (VIII) collagen signal peptide"
met_peptide	241..2412 /product="alpha-1 (VIII) collagen"
ALIGNMENT SCORES:	
Pred. NO.:	5.63e-06 Length: 2415
Score:	163.50 Matches: 45
Percent Similarity:	44.20% Conservative: 16
Best Local Similarity:	32.61% Mismatches: 48
Query Match:	22.96% Indels: 29
DB:	4 Gaps: 7
OS-09-997-610-2_COPY_18_149 (1-132) x RABCOLVIII (1-2415)	
OY 2	GLYPRORALAHISPROATRGPROFROGLUGLVALGLYPROROGIYALAProGIYLeu 21
Db 1840	GGCCAGGCTGCTGCTGCCAGCGAGGCCGCCAGGCTCCAGAGGGCTCCAGAGGGCTCCAGAGGCTGTG 1899
OY 22	-----ProGIYTYThGIYGLU---IleSerGIUmet----- 31
Db 1900	ATGGCCCCGACACAGCAGCACCAGGAGAGATCTGCACATATAGGGGGCTGGGGATTCAT 1959
OY 32	---ThrlYscYsPro-----CysProAspIle 39
Db 1960	GGGTTAAAACTCTCACGCTATGCGGCCCAAGAAAGGCAAGAACGGCGGGCCAGCCTAT 2019
OY 40	GLUATrSerAlaPheThrValIYLeuSerGIYLeuProLeuProPheIYSProlIe 59
Db 2020	GAGATGCTGCTGCTTACCGCGGAGCTGACGGCACCCTTCCCGCGGTGGGGCCCATTA 2079
OY 60	IlePheThrGIYValIleuTYrAsnAlaGlnArGIYsPleuIYsGIuAlaMetGIYValIlePhe 79

		::::::::::::::	:	:.....:	:::~::~
Bd	2080 AAGTTCACGACGACTGCCTTTACATGGCAGAACAAGCACTAACACC GCCCACGCCGCGCATCTTC				2139
Oy	80 AlacysarValProglYasntTYrTSerSerPheaspVaIguLeuhIsHiscylsLyS				99
Dd	2140 ACCTGCAGAGGTCCCGGGGCTTACTACTTGTCGAACACCGT-----CACTGCANA				2190
Oy	100 ---valsnlletPrleMetatrgLsglnlleulaAasnlyslgUgtlltle 116				
Dd	2191 GGGGGCAACGCTGGGGTTTT-----GCTCTGTTCAA GAACAACGACGCCGCTG				2235
RESULT 12					
BC013581					
LOCUS	BC013581	2506 bp	mRNA	linear	PRI 07-SEP-2001
DEFINITION	Homo sapiens, clone MGC:9568 IMAGE:3875911,				mRNA, complete cds.
ACCESSION	BC013581				
VERSION	BC013581.1 GI:15488903				
KEYWORDS	MGC.				
SOURCE	Homo sapiens.				
ORGANISM	Homo sapiens. Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Euarchonta; Primates; Catarrhini; Hominiidae; Homo. Strausberg, R.				
* REFERENCE	Direct Submission Submitted (04-SEP-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA				
AUTHORS	NIH-MGC Project URL: http://mgc.ncl.nih.gov				
JOURNAL	Contact: MGC help desk Email: cgaaps-femail.nih.gov Tissue Procurement: DCRD/DTP/Gazdar CDNA Library Preparation: Life Technologies, Inc. CDNA library Arrayed by: The I.M.A.G.E. Consortium (LNLT) DNA Sequencing by: Sequencing School at the Stanford Human Genome Center, Stanford University School of Medicine, Stanford, CA 94305 Web site: http://www.shgc.stanford.edu Contact: (Dickson, Mark) mcd@pxil.stanford.edu Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers, R. M.				
REMARK COMMENT					
CDS					
FEATURES					
source					
location/Qualifiers					
1..2506					
/organism="Homo sapiens"					
/db_xref="taxon:9606"					
/clone="MGC:9568 IMAGE:3875911"					
/tissue_type="lung, large cell carcinoma"					
/clone_lib="NIH_MGC_68"					
/lab_host="DH10B"					
/note="Vector: pCMV-Sport6"					
238..2472					
/codon_start=1					
/product="unknown (protein for MGC:9568)"					
/protein_id="AAHI3581.1"					
/db_xref="GI:15488904"					
/translation="MAVLPGPIOLDGLVTLSSTSRILIOAGAYVGIRKLPQIPQMPPDIPOYLGOOVPHMLPKDLAMKEKHRLIQGYEHLLPYMKELIDPARPMCKEEAVKKKEIKRPLASLRSGOGRGERGPSPRGSLRNHGIGRKGRPGGVGVKPEGMWGMRKGPCGMMGPCAKGEITGGEGEPMPGTRPGOPRPRLHPGIGRKGGGLPRDPGKRDRGRKGLCPGROGLRKPDKDFRMFGARCAVCNAPREMNRPPRCVCLPGYKRVETGTPGPQGUGLKPGAAGEBPQPPTICPVQGPSPPGIPGICPCODDIGIPQVPFPGVGGBOQLPGJLPGPPGLGIGKPFPGPKGBDRMGNGVPAALGPEREKGFPAIGAPEGPEPGLIEPIGPMKPPGAIFFGPGKGGCVIVGPGCGKPGERLDQFPKPGIAFLGVGSEPMRLPLPIDIGKEGAGOGKVGLGPGVGLDGKPBEPGIPGOSLOGPGIPGIGGSPPFIGPFIIGPKGEGRLPGRRPFCIGKRYVALYGAKGPGKPSALSPOGOCILGPPRPLGPPPPAVVMPTPPPQCYSIEDMDLGITGDVKPRIATYCARKKKNSGRAVEEMAFTEBELTAPEPPVAAPVFENKLTVNGRONYPDGIIFCEVPGVVYYFAHYHVHCKGGNWVALTFKNK					

SEQUENCING READ COVERAGE:sequencing is completed to a minimum standard of double strand coverage with a minimum of 2 clones and 2 reads with no ambiguities or 2 chemistries with a minimum of 2 clones and 3 reads with no ambiguities. If the sequence qualify for a region does not meet this standard, it will be indicated in the annotation as Low Coverage.

QUALITY OF INDIVIDUAL BASES. This sequence meets stringent quality standards - estimated error rate less than 1 per 10,000 bases. Reports of lowest quality individual bases and measures of base quality are listed below. Description of the metrics can be found at URL: <http://gc.bcm.tmc.edu:8088/quality.info/genbank.annotation.html>.

QUALSTAT-REPORT

FEATURES	Location/Qualifiers
source	1. .117000
	/organism="Homo sapiens"
	/db_xref="taxon:9606"
	/chromosome="3"
	/clone="RP11-383J23"
	/complement(1. .1988)
	/note="Overlaps bases 1. .1988 of clone AC024938"
misc_feature	/function="clone overlap"
repeat_region	2059. .2084
	/rpt_family="AT-rich"
repeat_region	2505. .2662
	/rpt_family="MIR"
repeat_region	complement(2805. .2881)
	/rpt_family="L2"
repeat_region	complement(2936. .3038)
	/rpt_family="L2"
repeat_region	complement(3024. .3183)
	/rpt_family="L2"
repeat_region	3300. .3464
	/rpt_family="MIR"
repeat_region	4436. .4779
	/rpt_family="MERLB"
repeat_region	5133. .5358
	/rpt_family="MER20"
repeat_region	complement(5611. .5770)
	/rpt_family="L1MA8"
repeat_region	complement(5809. .6018)
	/rpt_family="MIR"
repeat_region	8572. .8611
	/rpt_family="L2"
repeat_region	9064. .9344
	/rpt_family="Charlieb"
repeat_region	complement(9345. .9645)
	/rpt_family="Alu"
repeat_region	9646. .9790
	/rpt_family="Charlieb"
repeat_region	complement(10718. .10878)
	/rpt_family="L2"
repeat_region	11415. .11535
	/rpt_family="Charlieb2"
repeat_region	11869. .11731
	/rpt_family="CAGAG"
repeat_region	11973. .12104
	/rpt_family="Charlie2"
repeat_region	12285. .12440
	/rpt_family="MIR"
repeat_region	12802. .12940
	/rpt_family="MER5A"
repeat_region	13114. .13290
	/rpt_family="MER20"
repeat_region	15088. .15367
	/rpt_family="AluSC"
repeat_region	15368. .15395
	/rpt_family="TAAAA"
repeat_region	complement(16509. .16419)
	/rpt_family="MIR"

STS	19163. .19302	/standard_name="72547"	
STS	19297. .19426	/standard_name="171033"	
repeat_region	20164. .20228	/rpl_family="L2"	
STS	20770. .20907	/standard_name="43872"	
repeat_region	complement(22803. .23127)	/rpl_family="THELC"	
repeat_region	complement(23481. .23607)	/rpl_family="MER5A"	
repeat_region	23517. .23626	/rpl_family="MER5A"	
repeat_region	23651. .23770	/rpl_family="MER47B"	
repeat_region	complement(23771. .24106)	/rpl_family="MER1B"	
repeat_region	24107. .24307	/rpl_family="MER47B"	
repeat_region	complement(25342. .26197)	/rpl_family="L1PB1"	
repeat_region	27024. .27060	/rpl_family="MIR"	
repeat_region	27309. .27608	/rpl_family="A1u5g"	
Alignment Scores:			
Pred. No.:	0.000468	Length:	117000
Score:	162.50	Matches:	42
Percent Similarity:	42.75%	Conservative:	17
Best Local Similarity:	30.43%	Mismatches:	50
Query Match:	22.82%	Indels:	29
DB:	9	Gaps:	5
US-09-997-610-2_COPY_18_149 (1-132) x AC069222 (1-117000)			
QY	2	GLYPRProlAHISProAtgProBioLuGlUValGlyProProGlyAlaProGlyLeu	21
Db	72269	GGCCAGCCTGCCTTACGAGACCCCGAGCCCTCCAGGACCTCCAGAACCCCGACGCTGG	72210
QY	22	-----ProGlnTyrThrIleGluIle-----	28
Db	72209	ATGCGCCCTACACACACCCCGAGGAGATCTGCGAGATATGGCGGTGGAAATTGAT	72150
QY	29	-----serGluMetThrIlyScyProCyProaspIle	39
Db	72149	GGCGTGAACCCCCCATGCTACGGGGCTTAAGAAAGCAGCAATGAGGGCGACCTTAT	72090
QY	40	GLUArGserAlaPheThrValIlyLeuSerGlyIlyLeuProLeuProPheIlyProIle	59
Db	72089	GAGATGCTGCATTTACCGCGGAGCTAACCGCACCTTCCACCGGTGGGGCCCCAGTG	72030
QY	60	ILePheThrGlyValLeuTyrAsnAlaGlnAArgAspLeuIlyGluAlaMetGlyValPhe	79
Db	72029	AAGTTTAACAACAGCTGTATTAACGGCAGAGACGACTACAAACCCCGCAGACGATCTTC	71970
QY	80	AlAcysArGValProGlyAsnTyrTyrIserGserPheAspValGluLeuHisIscyIlyS	99
Db	71969	ACCTGTGAGGTCCCTGGGTGTACTACTTTCATCCACGTT-----CAGTCGAAG	71919
QY	100	---ValAsnIleTyrPleuMetArgIyGlnIleLeuAlaAsnIlyGluGluIle	116
Db	71918	GGGGGCAACGtGTGGGT-----GCTCTATTCAAGAACACAGACCCCGTG	71874
RESULT 14			
LOCUS	HS302D9/c	145880 bp	DNA linear PRI 12-DEC-1999
DEFINITION	Human DNA sequence from clone Rpl-302D9 on chromosome 22 Contains		
ACCESSION	282198		
VERSION	282198.2	GI:6572207	
KEYWORDS	HTG.		

SOURCE	Homo sapiens.
ORGANISM	Homo sapiens. Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. 1 (bases 1 to 145880)
REFERENCE	Bridgeman, A.
AUTHORS	Direct Submission
JOURNAL	Submitted (08-DEC-1999) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk requests: clonesrequest@sanger.ac.uk On Dec 13, 1999 this sequence version replaced gi:3164067.
COMMENT	During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above. This sequence has been finished according to sequence map criteria as follows. An attempt is made to resolve all sequencing problems, such as compressions and repeats, but not necessarily within known annotated human repeat sequence elements (e.g. Alu). Where the sequence is ambiguous, there is an annotation using the 'unsure' feature key. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em., EMBL; Sw., SWISSPROT; Tr., TREMBL; Wp., WORMEP; Information on the WORMEP database can be found at http://www.sanger.ac.uk/Projects/C_elegans/wormep This sequence was generated from part of bacterial clone contigs of human chromosome 22, constructed by the Sanger Centre Chromosome 22 Mapping Group. Further information can be found at http://www.sanger.ac.uk/HGP/Chr22 RPI-302D9 is from the library RPCI-1 constructed at the Roswell Park Cancer Institute by the group of Pieter de Jong. For further details see http://bacpac.med.buffalo.edu/ VECTOR: pCYPAC2 This sequence is the entire insert of clone RPI-302D9 The true left end of clone CTA-282F2 is at 69682 in this sequence. The true right end of clone CTA-415G2 is at 55167 in this sequence. location/Qualifiers 1..145880 /organism="Homo sapiens" /db_xref="taxon:9606" /chromosome="22" /clone="RPI-302D9" /clone_1ib-"RPCI-1" 188..245 /note="MER3 repeat: matches 144. .209 of consensus" 246..571 /note="Alusx repeat: matches 1. .312 of consensus" 572..759 /note="MER3 repeat: matches 1. .144 of consensus" 783..933 /note="MER5A repeat: matches 26. .187 of consensus" 1033..1336 /note="Alusp repeat: matches 1. .299 of consensus" 1450..1583 /note="MIR repeat: matches 24. .160 of consensus" 1687..1752 /note="L2 repeat: matches 2593. .2661 of consensus" 2350..2660 /note="Alusc repeat: matches 3. .309 of consensus" 2684..2981 /note="Alusg repeat: matches 2. .300 of consensus" 3323..3343 /note="MLTIE repeat: matches 116. .136 of consensus" 3344..3652 /note="Aluy repeat: matches 1. .309 of consensus" 3653..3928 /note="MLTIE repeat: matches 136. .359 of consensus" 3929..4278 /note="THEIB repeat: matches 3. .364 of consensus" 4279..4485


```

repeat_region      15304..15399
                    /note="MT1B repeat: matches 1..99 of consensus"
repeat_region      15490..15662
                    /note="AluSg1 repeat: matches 2..114 of consensus"
repeat_region      15669..15727
                    /note="MT1B repeat: matches 119..178 of consensus"
repeat_region      15728..16027
                    /note="AluSc repeat: matches 1..239 of consensus"
repeat_region      16028..16245
                    /note="MT1B repeat: matches 178..390 of consensus"
repeat_region      16546..16854
                    /note="AluY repeat: matches 1..300 of consensus"
repeat_region      18296..18323
                    /note="MSTA repeat: matches 2..29 of consensus"
repeat_region      18324..18392
                    /note="MER66-Internal repeat: matches 4919..4993 of
consensus"
repeat_region      18393..18712
                    /note="AluDb repeat: matches 1..311 of consensus"
repeat_region      18713..19133
                    /note="MER66-Internal repeat: matches 4548..4919 of
consensus"
misc_feature       complement(18872..19230)
                    /note="match: GSS: Em:A005063"
misc_feature       19251..19719
                    /note="match: GSS: Em:B14179"
repeat_region      19537..20290
                    /note="HERVH21 repeat: matches 4657..5784 of consensus"
repeat_region      20317..20382
                    /note="33 copies 2 mer ta 68 conserved"
repeat_region      20513..20666
                    /note="77 copies 2 mer tl 70 conserved"
repeat_region      20682..21008
                    /note="AluSg1 repeat: matches 1..306 of consensus"
repeat_region      21239..21553
                    /note="HUES-P3 repeat: matches 4410..4713 of consensus"
repeat_region      21882..22254
                    /note="THE1B repeat: matches 1..364 of consensus"
repeat_region      22302..22537
                    /note="MER66-Internal repeat: matches 2186..2417 of
consensus"
repeat_region      22538..22850
                    /note="AluSp repeat: matches 1..313 of consensus"
repeat_region      22851..23801
                    /note="MER66-Internal repeat: matches 1210..2186 of
consensus"
repeat_region      23905..23989
                    /note="MER66-Internal repeat: matches 3017..3102 of
consensus"

Alignment Scores:
Pred. No.:         0.000595      Length:      145880
Score:             162.50         Matches:    42
Percent Similarity: 50.89%        Conservative: 15
Best Local Similarity: 37.50%      Mismatches: 46
Query Match:       22.82%         Indels:    3
DB:                9              Gaps:        9

US-09-997-610-2_COPY_18_149 (1-132) x HS302D9 (1-145880)
QY 20 GtyleupProglIntyrThrglylIleSerGIuMetThrlyScySProCysProaspIle 39
    |||:|||||  ::  |||:|||||  ::  |||:|||||  |||:|||||
DB 90953 GGAATCCCGTTCGTTCTATGAGAA-----CAAGACCAAGCACTAGACAGCATATYT 90900
    |||:|||||  ::  |||:|||||  ::  |||:|||||  |||:|||||

QY 40 Gtlu-----ArgserAlaPheThrValIlyleuSerGIyIleuPro 53
    |||:|||||  |||:|||||  |||:|||||  |||:|||||  |||:|||||
DB 90899 GAAAAACGTTCAATCCCAACCAAGATTCGATTTGGTGTAACCTGATCGACCCCTCCCA 90840
    |||:|||||  |||:|||||  |||:|||||  |||:|||||  |||:|||||

QY 54 leupProPheIySProlleIlePheThrglyValIleuTyRAsnAlaGlnArGAspleuIyS 73
    ::|||:|||||  |||:|||||  |||:|||||  |||:|||||  |||:|||||
DB 90839 GAGCCCTTACGACCATCTGCTTCAAGCAAGCCCTTATATAACGATTCCTCCATTTCCAT 90780
    |||:|||||  |||:|||||  |||:|||||  |||:|||||  |||:|||||

QY 74 GtluAlaMetGIyAlaPheAlaIySarGValProGIyAsnTyTyrIySerPheaspVal 93
    |||:|||||  |||:|||||  |||:|||||  |||:|||||  |||:|||||

```

```

DB 90779 TTCTGCTGCAAGCATTCACCTCTACCAATCCCTAGTGTGACCATTTGTTGAATTT 90720
QY 94 GtluEuhIstIscysIySValAsnIleTrpleMetArGlySlnIle--leuAlaSn 112
    |||:|||||  |||:|||||  |||:|||||  |||:|||||  |||:|||||
DB 90719 GACCTGTTTCAGAGCTGTGCAATGTGCGATTTATATACGAATGCAATCCATATGAGAAAT 90660
QY 113 GtGluGIuIleSerIySngInIleGInGlu 124
    |||:|||||  |||:|||||  |||:|||||  |||:|||||  |||:|||||
DB 90659 AAACATGCTGTGCGCAAGCATGCGCTATGACAGCAAA 90624

RESULT 15
AC120598
LOCUS
DEFINITION
AC120598
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

REFERENCE
AUTHORS
Muzny,D.M., Adams,C., Adio-Oduola,B., Alt-osman,F.R., Allen,C.,
Alshrocks,S.L., AmaralLunge,H.C., Are,J.R., Ayala,M., Banks,T.,
Bardala,J., Benton,J., Bimaye,K., Blankenburg,K., Bonnin,D.,
Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
Buhay,C., Burgh,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
Douthwaite,K.J., Draper,H., Dusan-Rocha,S., Durbin,K.J.,
Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
Harris,C., Harris,K., Hart,M., Haylak,P., Hawes,A., Hernandez,J.,
Hernandez,O., Hodson,A., Hogue,M., Holloway,C., Hollins,B.,
Homai,P., Howard,S., Huber,J., Huiyk,S., Hume,J., Jackson,L.E.,
Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudan,S.,
Karlsdon,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
Li,J., Li,Z., Lichtarge,O., Liu,C., Liu,J., Liu,W., Louised,H.,
Lozodo,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
Massey,E., Mawhney,E., McLeod,M.P., Meador,M., Mel,G., Metzger,M.,
Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N.,
Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G.,
Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
Peterson,L., Pickens,R., Prims,E., Pul,L.L., Quiles,M., Ren,Y.,
Rives,M., Rojas,A., RojudoKani,I., Rolfe,M., Rolfe,S., Savery,G.,
Scherer,S., Scott,G., Shen,H., Shooshart,N., Sisson,I.,
Sodergren,E., Sonaikie,T., Sparks,A., Stanley,H., Stone,H.,
Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K.,
Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S.,
Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q.,
Wang,S., Ward-Moore,S., Warren,R., Washington,C., Wallington,S.,
Williams,G., Williamson,A., Wleczek,R., Wooden,S., Worley,K.,
Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
Weinstock,G. and Gibbs,R.
Direct Submission
unpublished
2 (bases 1 to 152354)
Worley,K.C.
Submitted 09-MAY-2002 Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 152354)
Worley,K.C.
Direct Submission

```

JOURNAL

COMMENT

Submitted (23-JUL-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
On Jul 18, 2002 this sequence version replaced g1:20564266.

----- Genome Center

Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>

Contact: hgsc-help@bcm.tmc.edu

----- Project Information

Center project name: GNMU

Center Clone name:

----- Summary Statistics

Sequencing vector: Plasmid;

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.990329

Consensus quality: 142151 bases at least Q40

Consensus quality: 147764 bases at least Q30

Consensus quality: 152093 bases at least Q20

* NOTE: Estimated insert size may differ from sequence length

* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).

* NOTE: This is a 'working draft' sequence. It currently

* consists of 33 contigs. The true order of the pieces

* is not known and their order in this sequence record is

* arbitrary. Gaps between the contigs are represented as

* runs of N, but the exact sizes of the gaps are unknown.

* This record will be updated with the finished sequence

* as soon as it is available and the accession number will

* be preserved.

1 2152: contig of 2152 bp in length
* 2153 2252: gap of unknown length
* 2253 4902: contig of 2650 bp in length
* 4903 5002: gap of unknown length
* 5003 7448: contig of 2446 bp in length
* 7449 7548: gap of unknown length
* 7549 10685: contig of 3137 bp in length
* 10686 10785: gap of unknown length
* 10786 13353: contig of 2568 bp in length
* 13354 13453: gap of unknown length
* 13454 16709: contig of 3256 bp in length
* 16710 16809: gap of unknown length
* 16810 20018: contig of 3209 bp in length
* 20019 20118: gap of unknown length
* 20119 23425: contig of 3307 bp in length
* 23426 23525: gap of unknown length
* 23526 26280: contig of 2755 bp in length
* 26281 26380: gap of unknown length
* 26381 29423: contig of 3043 bp in length
* 29424 29523: gap of unknown length
* 29524 32163: contig of 2640 bp in length
* 32164 32263: gap of unknown length
* 32264 36076: contig of 3813 bp in length
* 36077 36176: gap of unknown length
* 36177 39483: contig of 3307 bp in length
* 39484 39583: gap of unknown length
* 39584 43134: contig of 3551 bp in length
* 43135 43234: gap of unknown length
* 43235 46834: contig of 3600 bp in length
* 46835 46934: gap of unknown length
* 46935 52040: contig of 5106 bp in length
* 52041 52140: gap of unknown length
* 52141 57672: contig of 5532 bp in length
* 57673 57772: gap of unknown length
* 57773 62313: contig of 4541 bp in length
* 62314 62413: gap of unknown length
* 62414 66002: contig of 3589 bp in length
* 66003 66102: gap of unknown length
* 66103 72036: contig of 5934 bp in length
* 72037 72136: gap of unknown length
* 72137 77216: contig of 4980 bp in length
* 77217 80518: gap of unknown length
* 80518: contig of 3302 bp in length

* 80519 80618: gap of unknown length
* 80619 84810: contig of 4192 bp in length
* 84811 84910: gap of unknown length
* 84911 89660: contig of 4750 bp in length
* 89661 89760: gap of unknown length
* 89761 95933: contig of 6173 bp in length
* 95934 96033: gap of unknown length
* 96034 100988: contig of 4955 bp in length
* 100989 101088: gap of unknown length
* 101089 106799: contig of 5711 bp in length
* 106800 106899: gap of unknown length
* 106900 115241: contig of 8342 bp in length
* 115242 115341: gap of unknown length
* 115342 121457: contig of 6116 bp in length
* 121458 121557: gap of unknown length
* 121558 128525: contig of 6968 bp in length
* 128526 128625: gap of unknown length
* 128626 136102: contig of 7477 bp in length
* 136103 136202: gap of unknown length
* 136203 143456: contig of 7254 bp in length
* 143457 143556: gap of unknown length
* 143557 152354: contig of 8798 bp in length.
* Location/Qualifiers
1. 152354

FEATURES

source

1. /organism="Homo sapiens"

BASE COUNT 45796 a 28744 c 29399 g 44369 t 4046 others

ORIGIN

Alignment Scores:

Pred. No.: 0.00623 Length: 152354

Score: 162.50 Matches: 42

Percent Similarity: 42.75% Conservative: 17

Best Local Similarity: 30.43% Mismatches: 50

Query Match: 22.82% Indels: 29

DB: 2 Gaps: 5

US-09-997-610-2_COPY_18_149 (1-132) x AC120598 (1-152354)

Qy 2 GLYPROFROALHNSHSPROATGPROFROGLUVALGALYPROFROGLYALPROGLYLEU 21

Db 84076 GGCCAGCGCTGGCCCTCCAGACGCCAGGCCCTCCAGAGACTCCAGAGACCCCGACGCTGTG 84135

Qy 22 -----ProGlnTyrThrGlyGluLe----- 28

Db 84136 ATGCCCCCTACACACACACCCAGGAGATCTGCCAGATATGGGCTGGGAATTGAT 84195

Qy 29 -----SergLmethThryScysProGysProAspIle 39

Db 84196 GGCCTGAACCCGCCATGCCCTACGGGGCTAAGAAAGCAAGATGGAGGCCGACCTAT 84255

Qy 40 GLUARGSERALAPHETHRVALLYSLEUSERGLYLSLEUPROLEUPROPHELYPROILE 59

Db 84256 GAGATGCCGTGATTTACCGCGGACGTAAACCCACCTTCCACCGGGGGGGCCCGAGTG 84315

Qy 60 ILPEHETHRGlyValLeuTyrAsnAlaGlnArgAspLeuLysGluAlaMetGlyValPhe 79

Db 84316 AAGTTTAACAACAGCTGTATACGCAAGCAAGTAACAACCCGACAGCATCTTC 84375

Qy 80 ALAcysArgValLPRoGlyAsnTyrTyrSerSerGheAspValGluLeuHnshScyLys 99

Db 84376 ACCTGTAGAGGCTCGGTGCTACTACTTTCACAAACACGTT-----CACTGCAG 84426

Qy 100 ---ValAsnIleTrpLeuMetArgLysGlnIleLeuAlaAsnLysGluGluIle 116

Db 84427 GGGGCAACGtGTGGTT-----GCTGTATTCAAGAAACAAGCCCGTG 84471

Search completed: February 19, 2003, 23:01:51
Job time : 1969.21 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 19, 2003, 23:48:42 ; Search time 204.778 seconds
(without alignments)
6406.126 Million cell updates/sec

Title: US-09-997-610-1_COPY_56_136

Perfect score: 81

Sequence: 1 ggactccagcacacccccag.....cacaatcacagaggaata 81

Scoring table:

IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

EST:*
1: em_estba:*
2: em_estba:*
3: em_estba:*
4: em_estba:*
5: em_estba:*
6: em_estba:*
7: em_estba:*
8: em_estba:*
9: em_estba:*
10: em_estba:*
11: em_estba:*
12: em_estba:*
13: em_estba:*
14: em_estba:*
15: em_estba:*
16: em_estba:*
17: em_estba:*
18: em_estba:*
19: em_estba:*
20: em_estba:*
21: em_estba:*
22: em_estba:*
23: em_estba:*
24: em_estba:*
25: em_estba:*
26: em_estba:*
27: em_estba:*

Pred. NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	35.2	43.5	531	BE121287	UI-R-CAO-
2	35	43.2	461	BE121287	UI-R-CAO-
3	33.4	41.2	675	BE121287	UI-R-CAO-
4	33.4	41.2	894	BE121287	UI-R-CAO-
5	33.2	41.0	352	BE121287	UI-R-CAO-
6	33.2	41.0	360	BE121287	UI-R-CAO-

7	33.2	41.0	470	10	BE581111	BE581111	Kq42d08.y
8	33.2	41.0	632	13	Bj136441	Bj136441	
9	33.2	41.0	726	13	Bj108893	Bj108893	
10	33	40.7	613	10	AM179928	AM179928	
11	32.8	40.5	417	14	BM896730	BM896730	
12	32.8	40.5	698	14	BO548321	BO548321	
13	32.6	40.2	889	14	BO939199	BO939199	
14	32.4	40.0	494	14	BO258938	BO258938	
15	32.4	40.0	499	14	BM67732	BM67732	
16	32.4	40.0	690	10	BE201845	BE201845	
17	32	39.5	483	13	Bj122379	Bj122379	
18	32	39.5	552	13	Bj186371	Bj186371	
19	32	39.5	695	10	BE285930	BE285930	
20	32	39.5	776	13	Bj155467	Bj155467	
21	32	39.5	814	10	BE286914	BE286914	
22	31.8	39.3	269	9	AA56791	AA56791	
23	31.8	39.3	468	10	BE580165	BE580165	
24	31.8	39.3	483	10	BE579244	BE579244	
25	31.6	39.0	300	14	D70222	D70222	
26	31.6	39.0	360	14	C13375	C13375	
27	31.6	39.0	360	14	D69877	D69877	
28	31.6	39.0	420	10	BE580158	BE580158	
29	31.6	39.0	435	12	BF015196	BF015196	
30	31.6	39.0	438	12	BF015322	BF015322	
31	31.6	39.0	455	12	BF852009	BF852009	
32	31.6	39.0	458	12	BF015296	BF015296	
33	31.6	39.0	601	9	AU217866	AU217866	
34	31.6	39.0	645	13	Bj139557	Bj139557	
35	31.6	39.0	647	9	AU220646	AU220646	
36	31.6	39.0	666	13	Bj144812	Bj144812	
37	31.6	39.0	672	13	Bj138817	Bj138817	
38	31.6	39.0	672	13	Bj140783	Bj140783	
39	31.6	39.0	673	13	Bj144467	Bj144467	
40	31.6	39.0	674	13	Bj146794	Bj146794	
41	31.6	39.0	693	9	AU217833	AU217833	
42	31.6	39.0	707	13	Bj154926	Bj154926	
43	31.6	39.0	708	9	AU213793	AU213793	
44	31.6	39.0	708	13	Bj140880	Bj140880	
45	31.6	39.0	709	13	Bj147175	Bj147175	

ALIGNMENTS

RESULT 1	BE121287	531 bp	mRNA	linear	EST 13-JUN-2000
LOCUS	UI-R-CAO-baw-D-08-0-UI-s1	UI-R-CAO	Rattus norvegicus	CDNA clone	
DEFINITION	UI-R-CAO-baw-D-08-0-UI 3', mRNA sequence.				
ACCESSION	BE121287				
VERSION	BE121287.1	GI:8513392			
KEYWORDS	EST.				
SOURCE	Norway rat.				
ORGANISM	Rattus norvegicus				
REFERENCE	1 (bases 1 to 531)				
AUTHORS	Bonaldo,M.F., Lennon,G. and Soares,M.B.				
TITLE	Normalization and subtraction: two approaches to facilitate gene discovery				
JOURNAL	Genome Res. 6 (9), 791-806 (1996)				
MEDLINE	97044477				
COMMENT	Contact: Soares, MB Program for Rat Gene Discovery and Mapping University of Iowa 451 Eckstein Medical Research Building Iowa City, IA 52242, USA Tel: 319 335 8250 Fax: 319 335 9565 Email: msares@iuii.uow.edu The sequence contained an oligo-dT track that was present in the oligonucleotide that was used to prime the synthesis of first strand cDNA and therefore this may represent a bonafide poly A				

The Washington Univ. Nematode EST Project, 1999
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800

```

/organism="Onchocerca volvulus"
/strain="Kumba, Cameroons"
/db_xref="taxon:6282"
/clone="SMOV3MCAW56C07"
/clone_lib="Onchocerca volvulus molting L3 larva cDNA"

```


Matches	44	Conservative	0	Mismatches	18	Indels	0	Gaps	0
OY	1	GGACCTCCAGCACACCCAGGCCCCCAGAAAGATGGGGCTCTGCTGGTACACAGTTTA	60						
Db	101	GGACCTCTTGAGCACGTCGTCTGTGAGAAATGAGACCTCTGTCTGTCTCTT	160						
OY	61	CC	62						
Db	161	CC	162						
RESULT 6									
D69730			360 bp	mRNA	linear	EST 07-DEC-1995			
LOCUS									
DEFINITION									
CEIK073C7F									
clone yk73c7 5', mRNA sequence.									
D69730									
ACCESSION									
D69730.1									
VERSION									
KEYWORDS									
EST.									
Caenorhabditis elegans.									
Caenorhabditis elegans									
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea									
; Rhabditidae; Pelodirinae; Caenorhabditis.									
1 (bases 1 to 360)									
Kohara,Y., Mitsuiki,H., Nishigaki,A., Motonashi,T., Sugimoto,A. and									
Tohara,H.									
Toward an expression map of the C.elegans genome									
Unpublished (1994)									
Contact: Yuji Kohara									
Genome Biology Lab.									
National Institute of Genetics									
Yata 1111, Mishima, Shizuoka 411, Japan									
Tel: 81-559-81-6854									
Fax: 81-559-81-6855									
Email: ykohara@lab.nig.ac.jp.									
location/Qualifiers									
1. 360									
/organism="Caenorhabditis elegans"									
/strain="CB1489 him-8(el489)"									
/db_xref="taxon:6239"									
/clone="yk73c7"									
/clone_lib="Yuji Kohara unpublished CDNA"									
/sex="hermaphrodite, male"									
/tissue_type="whole animal"									
/dev_stage="varied"									
BASE COUNT	96 a	116 c	105 g	42 t	1	others			
ORIGIN									
Query Match									
Best Local Similarity	64.1%								
Matches	50;	Conservative	0;	Mismatches	28;	Indels	0;	Gaps	0;
OY	1	GGACCTCCAGCACACCCAGGCCCCCAGAAAGATGGGGCTCTGTCGTCACACAGTTTA	60						
Db	180	GGACCAACCAAGGCGCATCTCAGAGCCCCAGAGACAAAGGACCTTCAGAGAGCCCAAGATC	239						
OY	61	CCACATATACAGAGAA	78						
Db	240	CCAGACAAACCGAGGA	257						
RESULT 7									
LOCUS									
DEFINITION									
BE581111									
BE581111									
ACCESSION									
VERSION									
KEYWORDS									
EST.									
Strongyloides stercoralis.									
Strongyloides stercoralis									
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;									
Panagrolaimoidea; Strongyloidea; Strongyloides.									

REFERENCE		1 (bases 1 to 470)			
AUTHORS		McCarteer,J., Clifton,S., Chiappelli,B., Page,D., Martin,J., Wylie,T., Dante,M., Marra,M., Hillier,L., Kucaba,T., Theising,B., Powers,T., Gibbons,M., Ritter,E., Bennett,J., Franklin,C., Tsagarelisvili,R., Ronko,I., Kennedy,S., Maguire,L., Beck,C., Underwood,R., Steptoe ,M., Allen,M., Person,B., Swaller,T., Harvey,N., Schurk,R., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R. and Wilson,R.			
TITLE		The Washington Univ. Nematode EST Project, 1999			
JOURNAL		Unpublished (1999)			
COMMENT		Contact: McCarter JP The Washington Univ. Nematode EST Project, 1999 Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA Tel.: 314 286 1800 Fax: 314 286 1810 Email: est@wuston.wustl.edu The library was constructed by Dr. Thomas Nutman and colleagues of NIAID, NIH (nutman@nih.gov) DNA Sequencing by: Washington University Genome Sequencing Center St. Louis. High quality sequence stop: 416. Location/Qualifiers			
FEATURES	source	1..470			
		/organism="Strongyloides stercoralis" /strain="Rhabditiform larvae obtained from gerbils" /db_xref="taxon:6248" /clone_id="TBN95TM-SSR" /lab_host="XJ-1 Blue MRF" /note="Vector: Lambda Uni-ZAP XR (Stratagene); Site_1: EcoRI; Site_2: XhoI; mRNA was purified from 2 x 10E3 rhabditiform larvae which had been isolated from gerbils experimentally infected with larvae originally isolated from experimentally infected dogs. cDNA was constructed and, using adaptors, was cloned unidirectionally into the vector from the EcoRI site to the XhoI site. The library has an unamplified titer of 1 x 10E5 pfu/ml and an amplified, undiluted titer of 9 x 10E11 pfu/ml. The average insert size of the unamplified library is 675 bp (range, 100-1700)."			
BASE COUNT		147 a 132 c 117 g 74 t			
ORIGIN					
		Query Match 41.0%; Score 33.2; DB 10; Length 470; Best Local Similarity 64.1%; Pred. No. 9.1;			
Matches		50; Conservative 0; Mismatches 28; Indels 0; Gaps 0;			
OY	1	GGAGCTCGACGACACCCCGAGGCCCAAGAAGTGCGCCTGTGGCACAGATTGA 60			
Db	267	GGACCACCGACGACCTCCAGGGCCACCGAGATTTCCAGSAGAACCGACCGAGATTGA 326			
OY	61	CCACAATATACAGAGAGAA 78			
Db	327	CCAGGAAAAACGAGGTGAA 344			
RESULT 8					
LOCUS	Bj136441/c	632 bp mRNA linear EST 23-JAN-2002			
DEFINITION	Bj136441 unpublished oligo-capped cDNA library, C. elegans L1 stage				
ACCESSION	Bj136441				
VERSION	Bj136441.1	GI:18296598			
SOURCE	EST.				
ORGANISM	Caenorhabditis elegans.				
	Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabdilitida; Rhabditoidea ; Rhabditidae; Pelodetinae; Caenorhabditis.				
REFERENCE	Kohara,Y., Shin-i,T., Thierry-Mieg,J., Thierry-Mieg,D., Suzuki,Y.				
AUTHORS					
TITLE		A complementary view of the C.elegans genome			
JOURNAL		Unpublished (2002)			
COMMENT		Contact: Tadasu Shin-i			

Center For Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tshini@genes.nig.ac.jp.

FEATURES
source 1..632
Location/Qualifiers

/organism="Caenorhabditis elegans"
/strain="N2"
/db_xref="taxon:6239"
/clone="yk113b05"
/clone_lib="unpublished oligo-capped cDNA library, C. elegans L1 stage"
/sex="hermaphrodite"
/tissue_type="whole animal"
/dev_stage="L1"
/note="The AD-wrmcDNA library was generated with poly(A)+ RNA isolated from both hermaphrodite and male N2 worms of all larval stages, embryos, adults and dauers and the subsequent generation of cDNAs by poly(A) priming. The cDNAs were cloned into pPC86"

BASE COUNT 75 a 163 c 214 g 175 t 5 others
ORIGIN

Query Match 41.0%; Score 33.2; DB 13; Length 632;
Best Local Similarity 64.1%; Pred. No. 9.7;
Matches 50; Conservative 0; Mismatches 28; Indels 0; Gaps 0;

Oy 1 GGACCTCAGACACCCAGAGAGAGTGGGCTCTGTCGACACAGTTTA 60
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 496 GGACACACGACCATCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 437
Oy 61 CCACATATACAGAGAA 78
||| | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 436 CCAGACATCCGAGGA 419

RESULT 9
B108893
LOCUS B108893 726 bp mRNA linear EST 23-JAN-2002
DEFINITION B108893 unpublished oligo-capped cDNA library, C. elegans L1 stage
Caenorhabditis elegans cDNA clone yk113b05 5', mRNA sequence.
ACCESSION B108893
VERSION B108893.1 GI:18268921
KEYWORDS EST.
SOURCE Caenorhabditis elegans.
ORGANISM Caenorhabditis elegans
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea
1 (bases 1 to 726)
Rhabditidae; Peloderinae; Caenorhabditis.
Kohara, Y., Shinn, I., Thierry-Mieg, J., Thierry-Mieg, D., Suzuki, Y.
and Sugano, S.
A complementary view of the C.elegans genome
Unpublished (2002)
Contact: Tadasu Shinn-i
Center For Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tshini@genes.nig.ac.jp.

FEATURES
source 1..726
Location/Qualifiers

/organism="Caenorhabditis elegans"
/strain="N2"
/db_xref="taxon:6239"
/clone="yk113b05"
/clone_lib="unpublished oligo-capped cDNA library, C. elegans L1 stage"
/sex="hermaphrodite"
/tissue_type="whole animal"
/dev_stage="L1"

/note="The AD-wrmcDNA library was generated with poly(A)+ RNA isolated from both hermaphrodite and male N2 worms of all larval stages, embryos, adults and dauers and the subsequent generation of cDNAs by poly(A) priming. The cDNAs were cloned into pPC86"

BASE COUNT 189 a 219 c 192 g 125 t 1 others
ORIGIN

Query Match 41.0%; Score 33.2; DB 13; Length 726;
Best Local Similarity 64.1%; Pred. No. 9.7;
Matches 50; Conservative 0; Mismatches 28; Indels 0; Gaps 0;

Oy 1 GGACCTCAGACACCCAGAGAGAGTGGGCTCTGTCGACACAGTTTA 60
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 551 GGACACACGACCATCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 610
Oy 61 CCACATATACAGAGAA 78
||| | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 611 CCAGACATCCGAGGA 628

RESULT 10
AW179928
LOCUS AW179928 613 bp mRNA linear EST 16-NOV-1999
DEFINITION SMYD25CAU05E10SK Brugia malayi young adult day 25 cDNA
(SAM99MLW-BmyD25) Brugia malayi cDNA clone SMYD25CAU05E10 5', mRNA
sequence.
ACCESSION AW179928
VERSION AW179928.1 GI:6445965
KEYWORDS EST.
SOURCE Brugia malayi.
ORGANISM Brugia malayi.
Eukaryota; Metazoa; Nematoda; Chromadorea; Spirurida; Filarioidea;
Onchocercidae; Brugia.
1 (bases 1 to 613)
Genes expressed in young adult day 25 of Brugia malayi
Unpublished (1999)
Contact: Steven A. Williams
Molecular Parasitology
Smith College Department of Biological Sciences
Department of Biological Sciences, Clark Science Center, Smith
College, Northampton, MA, 01063, USA
Tel: 4135853826
Fax: 4135853786
Email: genome@smith.edu
Seq primer: pbluescript SK.
Location/Qualifiers

FEATURES
source 1..613
Location/Qualifiers

/organism="Brugia malayi"
/db_xref="taxon:6279"
/clone="SMYD25CAU05E10"
/clone_lib="Brugia malayi young adult day 25 cDNA
(SAM99MLW-BmyD25)"
/dev_stage="young adult, twenty five days after infection"
/lab_host="XLI-Blue MRF"
/note="Vector: Lambda Uni-ZAP XR; Site 1: Eco RI; Site 2:
Xho I; Lymphatic filarial nematode parasite of humans.
mRNA was prepared from young adult worms isolated from
the peritoneal cavity of jirds on day 25 after infection
and converted to double-stranded cDNA using reverse
transcriptase and oligo(dT) followed by RNase H and DNA
pol I. The library has 6.2 x 10⁵ independent recombinants
and the average insert size is approx. 110bp. The library
was constructed by Michelle Lizotte-Waniewski. The
library is available from Dr. S.A. Williams, email:
genome@neal.smith.edu."

BASE COUNT 181 a 139 c 175 g 118 t
ORIGIN

Query Match 40.7%; Score 33; DB 10; Length 613;
Best Local Similarity 63.0%; Pred. No. 11;
Matches 51; Conservative 0; Mismatches 30; Indels 0; Gaps 0;

RESULT	14
B0258938	
LOCUS	
DEFINITION	B0258938 494 bp mRNA linear EST 06-MAY-2002 taazze07.y1 zebrafish fin day3 regeneration Danilo cDNA clone 5910229.5' similar to TR:093486 O93486 ALPHA 3 TYPE I COLLAGEN ; contains PIR5.b2 MSRI repetitive element ;, mRNA sequence.
ACCESSION	B0258938
VERSION	B0258938.1 GI:20459702
KEYWORDS	EST.
SOURCE	zebrafish.
ORGANISM	Danio rerio Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes

RESULT 15	
BM967732	
LOCUS	499 bp mRNA linear EST 20-MAR-2002
DEFINITION	LM24HW0134 Bos taurus LM-24-HW CDNA library Bos taurus cDNA clone
ACCESSION	LM-24-HW-011-34 (5'), mRNA sequence.
VERSION	BM967732
KEYWORDS	BM967732.1 GI:19561919
SOURCE	EST.
ORGANISM	COW.
	Bos taurus
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
	Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidea;
	Bovidae; Bovinae; Bos.
REFERENCE	1 (bases 1 to 499)
AUTHORS	Yoon,D.H., Jang,Y.S., Kim,T.H., Park,E.W., Lee,H.K., Chung,E.R.,

Db 670 GGCCCTCCAGGACCCCAAGGGCCCCCAAGGAGAGTGGGCCCCCCCGGACCACCAGG 725

```

RESULT 2
US-09-029-348-20
: Sequence 20, Application US/09029348
: Patent No. 6171827
: GENERAL INFORMATION:
: APPLICANT: THE VICTORIA UNIVERSITY OF MANCHESTER
: TITLE OF INVENTION: NOVEL PROCOLLAGENS
: FILE REFERENCE: 0687857PUS LISTING
: CURRENT APPLICATION NUMBER: US/09/029,348
: CURRENT FILING DATE: 1998-05-07
: NUMBER OF SEQ ID NOS: 20
: SOFTWARE: patentIn Ver. 2.0
: SEQ ID NO 20
: LENGTH: 1881
: TYPE: DNA
: ORGANISM: Artificial Sequence
: FEATURE:
: OTHER INFORMATION: Description of Artificial Sequence: SEQUENCE
: OTHER INFORMATION: DERIVED FROM cDNA OF PROCOLLAGENS
US-09-029-348-20

```

[illegible]

```

RESULT 3
US-09-342-681C-7
Sequence 7, Application US/09342681C
Patent No. 6355782
GENERAL INFORMATION:
APPLICANT: Zonana et al.
TITLE OF INVENTION: Hypohydrotic ectodermal dysplasia genes and proteins
FILE REFERENCE: 52978
CURRENT APPLICATION NUMBER: US/09/342,681C
CURRENT FILING DATE: 1999-06-29
PRIOR APPLICATION NUMBER: 60/092,279
PRIOR FILING DATE: 1998-07-09
PRIOR APPLICATION NUMBER: 60/112,366
PRIOR FILING DATE: 1998-12-15
NUMBER OF SEQ ID NOS: 123
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 7
LENGTH: 821
TYPE: DNA
ORGANISM: Homo sapiens
US-09-342-681C-7

```

Query Match	36.3%	Score 29.4	DB 4	length 821
Best Local Similarity	60.8%	Pred. No. 0.62		
Matches 48	Conservative	0	Mismatches 31	Indels 0
QY 1	GGACCTTCAGACACCCGAGGCCCCAGAGAAGTGGGGCTCTGTGTCCACAGGTTTA	60		
Db 319	GGCCCTCCAGGACCCCGAGACTCTCAGGACCCCGAGGAGCCCGCAGGAATTCAGGGATT	378		
QY 61	CCACATATPACGAGGAAA	79		
Db 379	CCTGGAATTCGAGGACAA	397		

RESULT 4
US-09-342-681C-14

Sequence 14, Application US/09342681C
 Patent No. 6355782
 GENERAL INFORMATION:
 APPLICANT: Zonana et al.
 TITLE OF INVENTION: Hypohydrotic ectodermal dysplasia genes and proteins
 FILE REFERENCE: 5297B
 CURRENT APPLICATION NUMBER: US/09/342,681C
 CURRENT FILING DATE: 1999-06-29
 PRIOR APPLICATION NUMBER: 60/092,279
 PRIOR FILING DATE: 1998-07-09
 PRIOR APPLICATION NUMBER: 60/112,366
 PRIOR FILING DATE: 1998-12-15
 NUMBER OF SEQ ID NOS: 123
 SOFTWARE: PatentIn Ver. 2.1
 SEQ ID NO 14
 LENGTH: 1176
 TYPE: DNA
 ORGANISM: Homo sapiens
 FEATURE:
 NAME/KEY: CDS
 LOCATION: (1)..(1176)
 US-09-342-681C-14

	Query Match	Best Local Similarity	Matches	Conservative	Score 29.4;	DB 4;	length 1176;
					Pred. No. 0.68;		
					Mismatches 31;	Indels 0;	Gaps 0;
OY	1	GGACCTCCAGCAGACCCCGAGGCCCCAGAGAAAGTGGGGCCCTCGTGTACACAGATT	60				
Db	556	GGCCCTCCAGGACCCCGAGACCTCCAGAGACCCCGAGACCCCGAGATTCCAGGATT	615				
OY	61	CCACATATACAGAGAAA	79				
Db	616	CCTGGAATTCAGAGACAA	634				

```

RESULT 5
US-09-342-681C-1
Sequence 1, Application US/09342681C
Patent No. 635782
GENERAL INFORMATION:
APPLICANT: Zonana et al.
TITLE OF INVENTION: Hypohydrotic ectodermal dysplasia genes and proteins
FILE REFERENCE: 52978
CURRENT APPLICATION NUMBER: US/09/342,681C
CURRENT FILING DATE: 1999-06-29
PRIOR APPLICATION NUMBER: 60/092,279
PRIOR FILING DATE: 1998-07-09
PRIOR APPLICATION NUMBER: 60/112,366
PRIOR FILING DATE: 1998-12-15
NUMBER OF SEQ ID NOS: 123
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 1
LENGTH: 1574
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (242)..(1417)
US-09-342-681C-1

```

Query Match	36.3%	Score 29.4	DB 4	Length 1574
Best Local Similarity	60.8%	Pred. No. 0.74		
Matches 48	Conservative	0	Mismatches 31	Indels 0
Gaps				
Qy 1	GGACCTTCAGCAGACCCCGAGGCCCGAGAGAGGCGCCCTCGTGMCCACAGGTTTA	60		
Db 797	GGCCCTTCAGGAGCCCGCAGAGCCTTCAGAGACCCCGAGGACCCCGAGGAATTCAGGATT	856		
Qy 61	CCACAATATACGAGGAAA	79		
Db 857	CCTGGAATTCAGGAGACA	875		

Db 857 CCTGGA

```

RESULT 6
US-09-134-001C-635/c
: Sequence 635, Application US/09134001C
: Patent No. 6380370
:
GENERAL INFORMATION:
: APPLICANT: Lynn Doucellette-Stamm et al
: TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
: TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
: FILE REFERENCE: GTC-007
: CURRENT APPLICATION NUMBER: US/09/134,001C
: CURRENT FILING DATE: 1998-08-13
: PRIOR APPLICATION NUMBER: US 60/064,964
: PRIOR FILING DATE: 1997-11-08
: PRIOR APPLICATION NUMBER: US 60/055,779
: PRIOR FILING DATE: 1997-08-14
: NUMBER OF SEQ ID NOS: 5674
: SEQ ID NO 635
: LENGTH: 390
: TYPE: DNA
: ORGANISM: Staphylococcus epidermidis
: US-09-134-001C-635

```

Query Match	33.1%	Score 26.8	DB 4	Length 390
Best Local Similarity	59.0%	Pred. NO. 3.7		
Matches	46	Conservative	0	Mismatches 32
				Indels 0
				Gaps 0
QY	1	GGAGCTCCAGACACCCAGGCCCCCAGAGAAGTGGGGCTCTCTGGTGCACACAGTTTA	60	
Db	159	GGTACGCCAGCAGAACCGTAGTCGCCACAGAACCGTAGTTAAACCGCGAACCGAGTAGC	100	
QY	61	CCACAAATTAACAGAGAA	78	
Db	99	CCAGCAGAACCGAGTTAA	82	

```

RESULT 7
US-09-134-001C-647/c
: Sequence 647, Application US/09134001C
: Patent No. 6380370
:
: GENERAL INFORMATION:
: APPLICANT: Lynn Doucelte-Stamm et al
: TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
: TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
: FILE REFERENCE: GTC-007
: CURRENT APPLICATION NUMBER: US/09/134,001C
: CURRENT FILING DATE: 1998-08-13
: PRIOR APPLICATION NUMBER: US 60/064,964
: PRIOR FILING DATE: 1997-11-08
: PRIOR APPLICATION NUMBER: US 60/055,779
: PRIOR FILING DATE: 1997-08-14
: NUMBER OF SEQ ID NOS: 5674
: SEQ ID NO 647
: LENGTH: 513
:
: TYPE: DNA
: ORGANISM: Staphylococcus epidermidis
: US-09-134-001C-647

```

Query Match	33.1%	Score 26.8	DB 4	Length 513
Best Local Similarity	59.0%	Pred. No. 3.9		
Matches 46	Conservative 0	Mismatches 32	Indels 0	Gaps 0

OY	1	GGACCTCGAGACACCCGAGGCCCCGAGAAAGGCGCCTCTGTCGACACGTTTA	60
Db	298	GGTACGCCAGCAGAACCCGCTACGCCAGACGTAACCGAGCGAATCCAGGTACG	239
OY	61	CCACATTTACGAGGAA	78
Db	238	CCAGCAGAACCGGTTAA	221

RESULT 8
US-09-134-001C-726

```

: Sequence 726, Application US/09134001C
: Patent No. 6380370
: GENERAL INFORMATION:
: APPLICANT: Lynn Doucette-Stamm et al
: TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCC
: FILE REFERENCE: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
: CURRENT APPLICATION NUMBER: US/09/134,001C
: CURRENT FILING DATE: 1998-08-13
: PRIOR APPLICATION NUMBER: US 60/064,964
: PRIOR FILING DATE: 1997-11-08
: PRIOR APPLICATION NUMBER: US 60/055,779
: PRIOR FILING DATE: 1997-08-14
: NUMBER OF SEQ ID NOS: 5674
: SEQ ID NO 726
: LENGTH: 585
: TYPE: DNA
: ORGANISM: Staphylococcus epidermidis
US-09-134-001C-726

```

Query Match	33.1%	Score 26.8	DB 4	Length 585
Best Local Similarity	59.0%	Pred. No. 4.1		
Matches 46	Conservative	0	Mismatches 32	Indels 0
			Gaps	0
QY 1	GGACCTTCAGCAGCACCCGAGGCCCCAGAAAGTGGGGCTCTGTTGTCACACAGTTAA	60		
Db 270	GGTACGCCAGCAGAACCCAGTACGCCACAGAACCGATTAAACAGGGAACACAGGTACG	329		
QY 61	CCACAATATACAGAGAA	78		
Db 330	CCAGCAGAACCCAGTTAA	347		

```

RESULT 9
US-09-134-001C-624/c
: Sequence 624, Application US/09134001C
: Patent No. 6380370
: GENERAL INFORMATION:
: APPLICANT: Lynn Doucette-Stamm et al
: TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCC
: TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
: FILE REFERENCE: GTC-007
: CURRENT APPLICATION NUMBER: US/09/134,001C
: PRIORITY FILING DATE: 1998-08-13
: PRIOR APPLICATION NUMBER: US 60/064,964
: PRIORITY FILING DATE: 1997-11-08
: PRIOR APPLICATION NUMBER: US 60/055,779
: PRIORITY FILING DATE: 1997-08-14
: NUMBER OF SEQ ID NOS: 5674
: SEQ ID NO 624
: LENGTH: 810
: TYPE: DNA
: ORGANISM: Staphylococcus epidermidis
US-09-134-001C-624

```

Query Match	33.1%	Score 26.8	DB 4	length 810
Best Local Similarity	59.0%	Pred. No. 4.4		
Matches	46	Conservative	0	Mismatches 32; Indels 0; Gaps 0;
OY	1	GCACCTCCAGCAGACNCCCGAGCGCCCGCCAGCAAGAGTGGCGCCCTCCGTCGTCACACAGGTTA	60	
Db	266	GGTATGCCAGCGCAGAACCCAGGTATGCCAGCAGAACCAAGTAAACCGAGCGGACCAAGTACG	207	
OY	61	CCACATATACAGAGAA	78	
Db	206	CCAGCAGAACCAAGGTAAA	189	

```

RESULT 10
US-09-134-001C-693
; Sequence 693, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:

```

APPLICANT: Lynn Doucette-Stamm et al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: GTC-007
CURRENT APPLICATION NUMBER: US/09/134,001C
CURRENT FILING DATE: 1998-08-13
PRIOR FILING DATE: 1997-11-08
PRIOR APPLICATION NUMBER: US 60/055,779
PRIOR FILING DATE: 1997-08-14
NUMBER OF SEQ ID NOS: 5674
SEQ ID NO 693
LENGTH: 3552
TYPE: DNA
ORGANISM: Staphylococcus epidermidis
US-09-134-001C-693

Query Match 33.1%; Score 26.8; DB 4; Length 3552;
Best Local Similarity 59.0%; Pred. No. 6.5;
Matches 46; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

QY 1 GGACCTCCAGCACACCCAGCCCGAGAGAACTGGGCTCTGCTGCACAGTTTA 60
Db 3166 GGTACGCCAGAGAACCCAGGACCCAGACAGTAACCCAGGGAACCCAGTACG 3225

QY 61 CCACATATACAGAGAA 78
Db 3226 CCAGCAGAACCCAGTAA 3243

RESULT 11
5510466-3
PATENT NO. 5510466
APPLICANT: KREIGER, MONTY; KODAMA, TATSUHIKO
TITLE OF INVENTION: SCAVENGER RECEPTOR PROTEIN AND ANTIBODY
THERETO
NUMBER OF SEQUENCES: 12
CURRENT APPLICATION DATA:
APPLICATION NUMBER: 997,113
FILING DATE: 24-DEC-1992
APPLICATION NUMBER: 391,486
FILING DATE: 09-AUG-1989
APPLICATION NUMBER: 272,002
FILING DATE: 15-NOV-1988
SEQ ID NO: 3
LENGTH: 1588
5510466-3

Query Match 32.8%; Score 26.6; DB 6; Length 1588;
Best Local Similarity 58.0%; Pred. No. 6.1;
Matches 47; Conservative 0; Mismatches 34; Indels 0; Gaps 0;

QY 1 GGACCTCCAGCACACCCAGCCCGAGAGAACTGGGCTCTGCTGCACAGTTTA 60
Db 817 GGTCTCTCTGAGCTCCAGGTGAAAAAGAGATGAGAGCCCTCTGCAAAAATGTATA 876

QY 61 CCACATATACAGAGAAATA 81
Db 877 CCAGGCTTCCAGGTCTAATA 897

RESULT 12
5510466-1
PATENT NO. 5510466
APPLICANT: KREIGER, MONTY; KODAMA, TATSUHIKO
TITLE OF INVENTION: SCAVENGER RECEPTOR PROTEIN AND ANTIBODY
THERETO
NUMBER OF SEQUENCES: 12
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/307,400

FILING DATE: 16-SEP-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 997,113
FILING DATE: 24-DEC-1992
APPLICATION NUMBER: 391,486
FILING DATE: 09-AUG-1989
APPLICATION NUMBER: 272,002
FILING DATE: 15-NOV-1988
SEQ ID NO: 1
LENGTH: 1807
5510466-1

Query Match 32.8%; Score 26.6; DB 6; Length 1807;
Best Local Similarity 58.0%; Pred. No. 6.3;
Matches 47; Conservative 0; Mismatches 34; Indels 0; Gaps 0;

QY 1 GGACCTCCAGCACACCCAGCCCGAGAGAACTGGGCTCTGCTGCACAGTTTA 60
Db 849 GGTCTCTCTGAGCTCCAGGTGAAAAAGAGATGAGAGCCCTCTGCAAAAATGTATA 908

QY 61 CCACATATACAGAGAAATA 81
Db 909 CCAGGCTTCCAGGTCTAATA 929

RESULT 13
US-08-648-657-14
Sequence 14, Application US/08648657
Patent No. 5885813
GENERAL INFORMATION:
APPLICANT: Davis, Maria
APPLICANT: Moffett, R. Bruce
TITLE OF INVENTION: THERMOSTABLE DNA
TITLE OF INVENTION: POLYMERASES
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: Lyon & Lyon
STREET: 633 West Fifth Street
CITY: Los Angeles
STATE: California
COUNTRY: U.S.A.
ZIP: 90071-2066
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
MEDIUM TYPE: storage
COMPUTER: IBM Compatible
OPERATING SYSTEM: IBM P.C. DOS 5.0
SOFTWARE: Word Perfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/648,657
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
PRIOR APPLICATION DATA: including application one
PRIOR APPLICATION DATA: described below:
APPLICATION NUMBER: US 08/455,686
FILING DATE: May 31, 1995
ATTORNEY/AGENT INFORMATION:
NAME: Warburg, Richard J.
REGISTRATION NUMBER: 32,327
REFERENCE/DOCKET NUMBER: 219/304
TELECOMMUNICATION INFORMATION:
TELEPHONE: (213) 489-1600
TELEFAX: (213) 955-0440
TELEX: 67-3510
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 1686 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear

FEATURE:
NAME/KEY: FY4
LOCATION: 1..1683
US-08-648-657-14

Query Match 32.3%; Score 26.2; DB 2; Length 1686;
Best Local Similarity 63.5%; Pred. No. 8.4;
Matches 40; Conservative 0; Mismatches 23; Indels 0; Gaps 0;

OY 9 AGCACACCCAGCCCGGAGAGAGTGGGGCTCTGTGCACACAGTTTACCAATA 68
DB 1487 ACCTCTTCCCGCCGCGGAGATGGGGCCCGCATGCTCTCCAGTCCACGACGACC 1546

OY 69 TAC 71
DB 1547 TCC 1549

RESULT 14
PCT-US95-14418-3

Sequence 3, Application PC/TUS9514418
GENERAL INFORMATION:

APPLICANT:

APPLICANT:

TITLE OF INVENTION: DNA Encoding a Thermostable DNA Polymerase Enzyme

NUMBER OF SEQUENCES: 51

CORRESPONDENCE ADDRESS:

ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun

STREET: 6300 Sears Tower, 233 South Wacker Drive

CITY: Chicago

STATE: Illinois

COUNTRY: United States of America

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: PCT/US95/14418

FILING DATE:

CLASSIFICATION:

ATTORNEY/AGENT INFORMATION:

NAME: Gass, David A.

REGISTRATION NUMBER: 38,153

TELECOMMUNICATION INFORMATION:

TELEPHONE: 312/474-6300

TELEFAX: 312/474-0448

TELEX: 25-3856

INFORMATION FOR SEQ ID NO: 3:

SEQUENCE CHARACTERISTICS:

LENGTH: 1794 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

FEATURE:

NAME/KEY: CDS

LOCATION: 1..1794

PCT-US95-14418-3

Query Match 32.3%; Score 26.2; DB 5; Length 1794;

Best Local Similarity 63.5%; Pred. No. 8.5;
Matches 40; Conservative 0; Mismatches 23; Indels 0; Gaps 0;

OY 9 AGCACACCCAGCCCGGAGAGAGTGGGGCTCTGTGCACACAGTTTACCAATA 68
DB 1595 ACCTCTTCCCGCCGCGGAGATGGGGCCCGCATGCTCTCCAGTCCACGACGACC 1654

OY 69 TAC 71
DB 1655 TCC 1657

RESULT 15
PCT-US95-15327-3

Sequence 3, Application PC/TUS9515327
GENERAL INFORMATION:

APPLICANT:

TITLE OF INVENTION: Biologically Active Fragments of

TITLE OF INVENTION: Thermus Flavus DNA Polymerase

NUMBER OF SEQUENCES: 51

CORRESPONDENCE ADDRESS:

ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun

STREET: 6300 Sears Tower, 233 South Wacker Drive

CITY: Chicago

STATE: Illinois

COUNTRY: United States of America

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: PCT/US95/15327

FILING DATE:

CLASSIFICATION:

ATTORNEY/AGENT INFORMATION:

NAME: Gass, David A.

REGISTRATION NUMBER: 38,153

REFERENCE/DOCKET NUMBER: 28003/31716

TELECOMMUNICATION INFORMATION:

TELEPHONE: 312/474-6300

TELEFAX: 312/474-0448

TELEX: 25-3856

INFORMATION FOR SEQ ID NO: 3:

SEQUENCE CHARACTERISTICS:

LENGTH: 1794 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

FEATURE:

NAME/KEY: CDS

LOCATION: 1..1794

PCT-US95-15327-3

Query Match 32.3%; Score 26.2; DB 5; Length 1794;
Best Local Similarity 63.5%; Pred. No. 8.5;
Matches 40; Conservative 0; Mismatches 23; Indels 0; Gaps 0;

OY 9 AGCACACCCAGCCCGGAGAGAGTGGGGCTCTGTGCACACAGTTTACCAATA 68
DB 1595 ACCTCTTCCCGCCGCGGAGATGGGGCCCGCATGCTCTCCAGTCCACGACGACC 1654

OY 69 TAC 71
DB 1655 TCC 1657

Search completed: February 19, 2003, 22:58:45
Job time: 9.34727 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using SW model

Run on: February 19, 2003, 15:08:07 ; Search time 28.0336 Seconds
(without alignments)
6506.903 Million cell updates/sec

Title: US-09-997-610-1_COPY_56_136

Perfect score: 81

Sequence: 1 ggaccctccagcacaccacccag.....cacaatatacagagaata 81

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

N.Geneseq_101002:*

- 1: /SID52/gcgdata/geneeq/geneeqn-emb1/NA1980.DAT:*
- 2: /SID52/gcgdata/geneeq/geneeqn-emb1/NA1981.DAT:*
- 3: /SID52/gcgdata/geneeq/geneeqn-emb1/NA1982.DAT:*
- 4: /SID52/gcgdata/geneeq/geneeqn-emb1/NA1983.DAT:*
- 5: /SID52/gcgdata/geneeq/geneeqn-emb1/NA1984.DAT:*
- 6: /SID52/gcgdata/geneeq/geneeqn-emb1/NA1985.DAT:*
- 7: /SID52/gcgdata/geneeq/geneeqn-emb1/NA1986.DAT:*
- 8: /SID52/gcgdata/geneeq/geneeqn-emb1/NA1987.DAT:*
- 9: /SID52/gcgdata/geneeq/geneeqn-emb1/NA1988.DAT:*
- 10: /SID52/gcgdata/geneeq/geneeqn-emb1/NA1989.DAT:*
- 11: /SID52/gcgdata/geneeq/geneeqn-emb1/NA1990.DAT:*
- 12: /SID52/gcgdata/geneeq/geneeqn-emb1/NA1991.DAT:*
- 13: /SID52/gcgdata/geneeq/geneeqn-emb1/NA1992.DAT:*
- 14: /SID52/gcgdata/geneeq/geneeqn-emb1/NA1993.DAT:*
- 15: /SID52/gcgdata/geneeq/geneeqn-emb1/NA1994.DAT:*
- 16: /SID52/gcgdata/geneeq/geneeqn-emb1/NA1995.DAT:*
- 17: /SID52/gcgdata/geneeq/geneeqn-emb1/NA1996.DAT:*
- 18: /SID52/gcgdata/geneeq/geneeqn-emb1/NA1997.DAT:*
- 19: /SID52/gcgdata/geneeq/geneeqn-emb1/NA1998.DAT:*
- 20: /SID52/gcgdata/geneeq/geneeqn-emb1/NA1999.DAT:*
- 21: /SID52/gcgdata/geneeq/geneeqn-emb1/NA2000.DAT:*
- 22: /SID52/gcgdata/geneeq/geneeqn-emb1/NA2001A.DAT:*
- 23: /SID52/gcgdata/geneeq/geneeqn-emb1/NA2001B.DAT:*
- 24: /SID52/gcgdata/geneeq/geneeqn-emb1/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	39	48.1	1338	24	Human genset metab
2	32	39.5	6512	24	Mouse ischaemic co
3	32	39.5	61710	22	Human immune/haema
4	31.2	38.5	5010	24	Human Tumour Endot
5	31.2	38.5	6436	22	Human polynucleoti
6	31.2	38.5	51935	22	Human Immune/haema
7	30.4	37.5	549	24	cDNA #49 encoding
8	30.4	37.5	1317	22	Murine HSP47 inter
9	30.4	37.5	3394	18	Human alpha-1 col1

10	30.4	37.5	3394	20	AAx78379	Human alpha (XVII
11	30.4	37.5	3394	24	ABN95680	Gene #2178 used to
12	30.2	37.3	5467	22	AAH98343	Human EST-derived
13	30.2	37.3	5468	22	AAH98411	Human EST-derived
14	30	37.0	707	22	AA544890	Human contig polyn
15	30	37.0	2686	22	AA527013	CDNA encoding nove
16	30	37.0	3321	22	AA544718	Human full-length
17	29.8	36.8	561	24	ABK44107	CDNA #47 encoding
18	29.8	36.8	617	24	ABK44111	CDNA #51 encoding
19	29.8	36.8	648	24	ABK44126	CDNA #66 encoding
20	29.8	36.8	653	24	ABK44113	CDNA #53 encoding
21	29.8	36.8	655	24	ABK44128	CDNA #68 encoding
22	29.8	36.8	656	24	ABK44129	CDNA #69 encoding
23	29.8	36.8	658	24	ABK44131	CDNA #71 encoding
24	29.8	36.8	659	24	ABK44112	CDNA #52 encoding
25	29.8	36.8	780	24	ABK44065	CDNA #5 encoding h
26	29.8	36.8	823	24	ABK44088	CDNA #28 encoding h
27	29.8	36.8	994	22	AAE44902	Human breast cance
28	29.8	36.8	1881	18	AAE59892	Coding sequence fo
29	29.8	36.8	4428	22	AAE06574	Bovine alpha(III)
30	29.8	36.8	4428	22	AAE06575	Porcine alpha(III)
31	29.8	36.8	4428	22	AAE06578	Porcine alpha(III)
32	29.8	36.8	5460	17	AAE16508	Vector pAC3A1 cont
33	29.8	36.8	5460	22	ABAE83117	Collagen type III
34	29.8	36.8	5460	24	ABAE92101	Human Tumour Endot
35	29.8	36.8	5466	23	AAE79378	DNA encoding novel
36	29.8	36.8	9287	24	ABK64501	Human benign prost
37	29.6	36.5	1619	22	AAE27365	CDNA encoding nove
38	29.6	36.5	1619	22	AAE34830	CDNA encoding nove
39	29.6	36.5	6674	22	AAE58233	Human polynucleoti
40	29.4	36.3	821	24	ABE51013	Human ED41-II exon
41	29.4	36.3	1176	24	ABE51020	Human ED41-II open
42	29.4	36.3	1574	24	ABE51009	Human ED41-II enco
43	29.4	36.3	4062	22	AAH33347	Human colour cancer
44	29.4	36.3	9192	24	ABE92117	Human Tumour Endot
45	29.4	36.3	9483	22	AAE44690	Human full-length

ALIGNMENTS

RESULT 1	AAE44066	AAE44066 standard; cDNA: 1338 BP.
ID	AAE44066	
AC	AAE44066;	
XX		
DT	27-SEP-2002 (first entry)	
XX		
DE	Human genset metabolic gene (GMC-9) cDNA sequence.	
XX		
KW	Human; gene; ss; gene therapy; genset metabolic gene; GMC-7A; GMC-7B;	
KW	GMC-8; GMC-9; GMC-10; GMC-11; metabolic-related disorder; obesity;	
KW	Impaired glucose tolerance; insulin resistance; Syndrome X;	
KW	Type II diabetes; hyperlipidaemia; atherosclerosis; hypertension;	
KW	heart disease; cardiac insufficiency; coronary insufficiency;	
KW	high blood pressure; insulin sensitizer;	
XX	non-insulin dependent diabetes mellitus.	
XX		
OS	Homo sapiens.	
XX		
FT	Key	Location/Qualifiers
FT	CDS	1..1338
FT		/*tag= a
FT		/partial
FT		/product= "Human GMC-9 protein"
FT		/note= "No stop codon is given"
PN	WO200255694-A2.	
XX		
PD	18-JUL-2002.	
XX		
PF	15-JAN-2002; 2002WO-1B01215.	

```

XX 16-JAN-2001; 2001US-262235P.
PR
XX (GENSET ) GENSET.
PA
XX
XX Erickson MR, Bour BA, Bihain B, Tanaka H;
PI
XX WPI: 2002-557821/59.
DR
XX P-PSDB; AAO15423.
DR
XX
XX Treating or preventing a metabolic-related disease or disorder, e.g.
PT obesity, impaired glucose tolerance, insulin resistance, Syndrome X, or
PT Type II diabetes, comprises administering Genset Metabolic genes -
PS
XX
XX Disclosure: Page 122-124; 128pp; English.
XX
XX The invention comprises the amino acid and coding sequences of six human
CC genset metabolic genes (GMG-7A, GMG-7B, GMG-8, GMG-9, GMG-10 and GMG-11).
CC The GMG DNA and protein sequences of the invention are useful for
CC treating or preventing metabolic-related disorders, such as: obesity;
CC impaired glucose tolerance; insulin resistance; Syndrome X; Type II
CC diabetes; hyperlipidaemia; atherosclerosis; hypertension; and heart
CC diseases (e.g. cardiac insufficiency, coronary insufficiency or high
CC blood pressure). The GMG DNA and protein sequences of the invention may
CC also be used as insulin sensitizers - for improving insulin sensitivity
CC in persons with non-insulin dependent diabetes mellitus. The present cDNA
CC sequence encodes the human GMG-9 protein.
XX
XX
SQ Sequence 1338 BP; 359 A; 288 C; 350 G; 341 T; 0 other;

Query Match 48.1%; Score 39; DB 24; Length 1338;
Best Local Similarity 89.4%; Pred. NO. 0.0035;
Matches 42; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 35 TGGGCGCTCCTGTCGACAGGTTTACGCACATATACAGGAGAAATA 81
Db ||| ||||||| ||||||| ||||||| ||||||| |||||||
50 TGGATGTCTCTGTGTCCCGAGGTTTACCATATACAGGAGAAATA 96

RESULT 2
AB199819
ID AB199819 standard; cDNA: 6512 BP.
XX
XX AB199819;
AC
XX
XX 07-MAR-2002 (first entry)
DT
XX
XX Mouse ischaemic condition related cDNA sequence SEQ ID NO:932.
DE
XX
XX Mouse; ischaemia; compressive ischaemia; occlusive ischaemia;
KW vasospastic ischaemia; ischaemic condition; ischaemic disease; ss.
KW
XX
XX Mus musculus.
OS
XX
XX WO200188188-A2.
PN
XX
XX 22-NOV-2001.
PD
XX
XX 18-MAY-2001; 2001WO-JP04192.
PE
XX
XX 18-MAY-2000; 2000JP-0145977.
PR
XX
XX (UYNI-) UNITV NIHON SCHOOL JURIDICAL PERSON.
PA
XX
XX Ishikawa K, Asai S, Takahashi Y, Nagata T, Ishii Y;
PI
XX
XX WPI: 2002-034733/04.
XX
XX P-PSDB; ABB57334.
DR
XX
XX Examining the ischemic condition (e.g. occlusive ischemia) by measuring
PT expression levels of particular genes defined in the specification or
PT by determining the expression profile of a gene group comprising these
PT genes -

```

XX	Claim 2; Page 2340-2352; 2690pp; English.
XX	
CC	The present invention describes a method for examining ischemic
CC	conditions, comprising measuring the expression levels of particular
CC	genes (I) in a test sample or determining the expression profile of a
CC	gene group in the sample comprising genes selected from (I). The method
CC	is useful for examining the ischemic condition (e.g. compressive
CC	ischemia, occlusive ischemia or vasospastic ischemia) by measuring
CC	expression levels of particular genes (AB199202 to AB199312, encoding
CC	the protein sequences in AB857020 to AB857374) or by determining the
CC	expression profile of a gene group comprising these genes. The
CC	expression levels or expression profiles produced by these genes are
CC	used as an indicator when screening for ischemic condition-improving
CC	drugs or therapeutics for ischemic diseases. AB199313 and AB199314
CC	represent PCR primers for a mouse ischemic condition related sequence,
CC	which are used in the exemplification of the present invention.
XX	
SO	Sequence 6512 BP; 1553 A; 1798 C; 1950 G; 1211 T; 0 other:
	Query Match 39.5%; Score 32; DB 24; Length 6512;
	Best Local Similarity 62.5%; Pred. No. 1;
	Matches 50; Conservative 0; Mismatches 30; Indels 0; Gaps 0;
QY	1 GGACCTCCAGCAGCACACCAGCCGCCACAGAGAAGTGGGCTCTGTGTCACACAGTTTA 60
DB	2694 GGACAGCTGAGCGGCTCCCTGGCCCTCTGACAGCAGGGGACACCTGGAATTCAGAGGTTT 2753
QY	61 CCACAAATATACAGGAGAAAT 80
DB	2754 CCAGGTCTTAAGAGTGAAAT 2773
	RESULT 3
	AAK83782/C
ID	AAK83782 standard; DNA; 61710 BP.
XX	
AC	AAK83782;
XX	
DT	07-NOV-2001 (first entry)
XX	
DE	Human immune/haematopoietic antigen genomic sequence SEQ ID NO:38594.
XX	
KW	Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW	cytostatic; gene therapy; vaccine; metastasis; ds.
XX	
OS	Homo sapiens.
XX	
PN	WO200157182-A2.
XX	
PD	09-AUG-2001.
XX	
FE	17-JAN-2001; 2001WO-US01354.
XX	
PR	31-JAN-2000; 2000US-0179065.
PR	04-FEB-2000; 2000US-0180628.
PR	24-FEB-2000; 2000US-0184664.
PR	02-MAR-2000; 2000US-0186350.
PR	16-MAR-2000; 2000US-0189874.
PR	17-MAR-2000; 2000US-0190076.
PR	18-APR-2000; 2000US-0198123.
PR	19-MAY-2000; 2000US-0205515.
PR	07-JUN-2000; 2000US-0209467.
PR	28-JUN-2000; 2000US-0214886.
PR	30-JUN-2000; 2000US-0215135.
PR	07-JUL-2000; 2000US-0216647.
PR	11-JUL-2000; 2000US-0216880.
PR	11-JUL-2000; 2000US-0217487.
PR	11-JUL-2000; 2000US-0218290.
PR	14-JUL-2000; 2000US-0220963.
PR	26-JUL-2000; 2000US-0220964.
PR	14-AUG-2000; 2000US-0224518.
PR	

PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-022547.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 23-AUG-2000; 2000US-0227182.
PR 30-AUG-2000; 2000US-0227009.
PR 01-SEP-2000; 2000US-0228287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239335.
PR 13-OCT-2000; 2000US-0239337.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.

PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246538.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251038.
PR 05-DEC-2000; 2000US-0251188.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251858.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251899.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.

(HUMA-) HUMAN GENOME SCI INC.
Rosen CA, Barash SC, Ruben SM;
WPI; 2001-483426/52.

Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
useful for preventing, diagnosing and/or treating cancers and
metastasis -

Disclosure: SEQ ID NO 38594; 3071pp + Sequence Listing: English.

AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic
activity, and can be used in gene therapy and vaccine production. (I)
proteins and polynucleotides may be used in the prevention, diagnosis and
treatment of diseases associated with inappropriate (I) expression. For
example, they may be used to treat disorders associated with decreased
expression by rectifying mutations or deletions in a patient's genome
that affect the activity of (I) by expressing inactive proteins or to
supplement the patients own production of (I). Additionally, (I)
polynucleotides may be used to produce the secreted (I), by inserting
the nucleic acids into a host cell and culturing the cell to express the
protein. (I) proteins and polynucleotides may be used to prevent,
diagnose and treat immune/haematopoietic-related diseases, especially
cancers and cancer metastases of haematopoietic-derived cells. AAK64703
to AAK87694 represent human immune/haematopoietic antigen genomic
sequences from the present invention. AAK54942 to AAK54950 and AAM82169

CC represent sequences used in the exemplification of the present invention.
XX Sequence 61710 BP; 11560 A; 18542 C; 18148 G; 13460 T; 0 other;
SQ Best Local Similarity 39.5%; Score 32; DB 22; Length 61710;
Query Match Best Local Similarity 65.3%; Pred. No. 1.8;
Matches 47; Conservative 0; Mismatches 25; Indels 0; Gaps 0;
OY 1 GGACCTCCAGCACACCCAGGCCCCAGAGAGAGTGGGGCCTCGTGACACAGTTTA 60
DB 44831 GGCCCTCCAGGACCAAGGGCGCCAAAGAGAGAGTGGGGCCCGCCGACACACAGGTGAG 44772
OY 61 CCACATATATACA 72
DB 44771 CAACTCTGGACA 44760
RESULT 4
ABL92114
ID ABL92114 standard; cDNA; 5010 BP.
XX ABL92114;
AC ABL92114;
XX 30-MAY-2002 (first entry)
DT 30-MAY-2002 (first entry)
DE Human Tumour Endothelial Marker polynucleotide SEQ ID NO 251.
XX Human Tumour Endothelial Marker polynucleotide SEQ ID NO 251.
XX Human; mouse; rat; TEM; tumour endothelial marker; NEM; PEM; cytotstatic;
KW normal endothelial marker; pan-endothelial marker; immunostimulant;
KM antiangiogenic; tumour; neovascularization; vascularised tumour;
KW polycystic kidney disease; diabetes; retinopathy; rheumatoid arthritis;
KM psoriasis; gene; ss.
XX Homo sapiens.
OS Homo sapiens.
XX WO200210217-A2.
PN WO200210217-A2.
XX 07-FEB-2002.
PD 07-FEB-2002.
XX 01-AUG-2001; 2001WO-0524031.
PF 01-AUG-2001; 2000US-222599P.
XX 02-AUG-2000; 2000US-222599P.
PR 11-AUG-2000; 2000US-224360P.
XX 11-APR-2001; 2001US-282850P.
PR 11-APR-2001; 2001US-282850P.
XX (UYJO) UNITV JOHNS HOPKINS.
PA (UYJO) UNITV JOHNS HOPKINS.
XX St Croix B. Kinzler KW, Vogelstein B;
PI St Croix B. Kinzler KW, Vogelstein B;
XX WPI; 2002-291856/33.
DR P-PSDB; ABB90760.
XX An isolated molecule comprising an antibody variable region which
PT specifically binds to an extracellular domain of a tumor endothelial
marker (TEM) protein, useful for inhibiting tumor growth -
XX Claim 58; Page 245-246; 331pp; English.
PS The invention relates to an isolated molecule comprising an antibody
XX variable region which specifically binds to an extracellular domain of a
CC tumour endothelial marker (TEM) protein selected from ABB90732, ABB90740,
CC ABB90749, ABB90750 and ABB90769. The antibodies which bind to TEM
CC proteins have cytostatic, immunostimulant and antiangiogenic activity.
CC They are useful for inhibiting tumour growth, neovascularization in
CC subjects bearing a vascularised tumour, polycystic kidney disease,
CC diabetic retinopathy, rheumatoid arthritis and psoriasis. Human, mouse
CC and rat TEM genes and the encoded proteins (ABL92075-ABL92141 and
CC ABB90721-ABB90789) are disclosed, as are marker oligonucleotide
CC sequences: tumour endothelial markers (TEM) ABL91996-ABL92041 and
CC ABL92143-ABL92191; normal endothelial markers (NEM) ABL92042-ABL92074;
CC and pan-endothelial markers (PEM) ABL91903-ABL91995.
XX Sequence 5010 BP; 1157 A; 1354 C; 1658 G; 841 T; 0 other;

Query Match 38.5%; Score 31.2; DB 24; Length 5010;
Best Local Similarity 66.2%; Pred. No. 1.8;
Matches 45; Conservative 0; Mismatches 23; Indels 0; Gaps 0;
OY 13 CACCCAGGCCCCAGAGAGTGGGGCCCGCTGGTGACACAGTTTTCACCATATACA 72
DB 1303 CAGCCCGACCTCCAGGTGACAGGGTCTCTGGAATTTCAGGGCAGCCAGATTATTA 1362
OY 73 GGAGAAAT 80
DB 1363 GGCGAAAT 1370
RESULT 5
AA160019
ID AA160019 standard; cDNA; 6436 BP.
XX AA160019;
AC AA160019;
XX 22-OCT-2001 (first entry)
DT 22-OCT-2001 (first entry)
DE Human polynucleotide SEQ ID NO 4008.
XX Human polynucleotide SEQ ID NO 4008.
XX Human; neotropic; immunosuppressant; cytotstatic; gene therapy; cancer;
KW peripheral nervous system; neuropathy; central nervous system; CNS;
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW chemokine; thrombolytic; drug screening; arthritis; inflammation;
KW leukaemia; ss.
XX Homo sapiens.
OS Homo sapiens.
XX WO200153312-A1.
PN WO200153312-A1.
XX 26-JUL-2001.
PD 26-JUL-2001.
XX 26-DEC-2000; 2000WO-US34263.
PF 26-DEC-2000; 2000US-0488725.
XX 21-JAN-2000; 2000US-0488725.
PR 25-APR-2000; 2000US-0552317.
XX 09-JUL-2000; 2000US-0598042.
PR 19-JUL-2000; 2000US-0620312.
XX 03-AUG-2000; 2000US-0653450.
PR 14-SEP-2000; 2000US-0662191.
XX 19-OCT-2000; 2000US-0693036.
PR 29-NOV-2000; 2000US-0727344.
XX (HYSE-) HYSEQ INC.
PA (HYSE-) HYSEQ INC.
XX Tang YT, Liu C, Asundi V, Chan R, Ma Y, Qian XB, Ren F, Wang D;
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
XX Zhao QA, Zhou P, Goodrich R, Drmanac RT;
PI WPI; 2001-442253/47.
DR WPI; 2001-442253/47.
XX P-PSDB; AAM40863.
XX Novel nucleic acids and polypeptides, useful for treating disorders
PT such as central nervous system injuries -
XX Claim 1; SEQ ID NO 4008; 10078pp; English.
PS The invention relates to human nucleic acids (AA157798-AA161369) and
XX the encoded polypeptides (AAM38642-AAM42213) with neotropic;
CC immunosuppressant and cytotstatic activity. The polynucleotides are useful
CC in gene therapy. A composition containing a polypeptide or polynucleotide
CC of the invention may be used to treat diseases of the peripheral nervous
CC system, such as peripheral nervous injuries, peripheral neuropathy and
CC localised neuropathies and central nervous system diseases, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC utilisation of the activities such as: immune system suppression,
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic

PR	17-NOV-2000.	2000US-0249207.
PR	17-NOV-2000.	2000US-0249208.
PR	17-NOV-2000.	2000US-0249209.
PR	17-NOV-2000.	2000US-0249210.
PR	17-NOV-2000.	2000US-0249211.
PR	17-NOV-2000.	2000US-0249212.
PR	17-NOV-2000.	2000US-0249213.
PR	17-NOV-2000.	2000US-0249214.
PR	17-NOV-2000.	2000US-0249215.
PR	17-NOV-2000.	2000US-0249216.
PR	17-NOV-2000.	2000US-0249217.
PR	17-NOV-2000.	2000US-0249218.
PR	17-NOV-2000.	2000US-0249244.
PR	17-NOV-2000.	2000US-0249245.
PR	17-NOV-2000.	2000US-0249264.
PR	17-NOV-2000.	2000US-0249265.
PR	17-NOV-2000.	2000US-0249297.
PR	17-NOV-2000.	2000US-0249299.
PR	17-NOV-2000.	2000US-0249300.
PR	01-DEC-2000.	2000US-0250160.
PR	01-DEC-2000.	2000US-0250391.
PR	05-DEC-2000.	2000US-0251030.
PR	05-DEC-2000.	2000US-0251988.
PR	05-DEC-2000.	2000US-0256719.
PR	06-DEC-2000.	2000US-0251479.
PR	08-DEC-2000.	2000US-0251856.
PR	08-DEC-2000.	2000US-0251868.
PR	08-DEC-2000.	2000US-0251869.
PR	08-DEC-2000.	2000US-0251989.
PR	08-DEC-2000.	2000US-0251990.
PR	11-DEC-2000.	2000US-0254097.
PR	05-JAN-2001.	2001US-0259678.

PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Rosen CA, Barash SC, Ruben SM.
 XX
 DR WPI: 2001-483426/52.
 XX
 PT Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
 PT useful for preventing, diagnosing and/or treating cancers and
 XX metastasis -
 PS
 PS Disclosure; SEQ ID NO 30695; 3071bp + Sequence Listing; English.
 XX
 CC AAK54951 to AAK64702 encode the human immune/hematopoietic antigen (I)
 CC amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic
 CC activity, and can be used in gene therapy and vaccine production. (I)
 CC proteins and polynucleotides may be used in the prevention, diagnosis and
 CC treatment of diseases associated with inappropriate (I) expression. For
 CC example, they may be used to treat disorders associated with decreased
 CC expression by rectifying mutations or deletions in a patient's genome
 CC that affect the activity of (I) by expressing inactive proteins or to
 CC supplement the patients own production of (I). Additionally, (I)
 CC polynucleotides may be used to produce the secreted (I), by inserting
 CC the nucleic acids into a host cell and culturing the cell to express the
 CC protein. (I) proteins and polynucleotides may be used to prevent,
 CC diagnose and treat immune/hematopoietic-related diseases, especially
 CC cancers and cancer metastases of hematopoietic-derived cells. AAK64703
 CC to AAK87694 represent human immune/hematopoietic antigen genomic
 CC sequences from the present invention. AAK54942 to AAK54950 and AAM82169
 CC represent sequences used in the exemplification of the present invention.
 XX
 XQ Sequence 51935 BP; 14418 A; 12317 C; 11675 G; 13525 T; 0 other;

[illegible]

QY 73 GGAGGAAT 80
|||
Db 46084 GGCGAAT 46077

RESULT 7
ABK44109
ID ABK44109 standard; cDNA; 549 BP.

AC	ABK44109;
XX	
DT	21-MAY-2002 (first entry)

DE CDNA #49 encoding human pancreatic tumour protein.

KW Human; pancreatic tumour protein; immune response; pancreatic cancer development of cancer; cancer progression; cytostatic; gene; ss.

05 Homo sapiens

PN WO200212331-A2.

PD 14-FEB-2002.

PF 06-AUG-2001; 2001WO-US24619.

PR 07-AUG-2000; 2000US-223130P.

PR 15-MAY-2001; 2001US-291201P.

PA (CORI-) CORIXA CORP.

PI Pyle RA, Xu J, Kalos MD;

DR WPI; 2002-241741/29.

PT Novel polynucleotide encoding pancreatic tumour polypeptides, useful in

PT cancers -

PS Claim 1; page 126; 167pp; English.

CC The present invention relates to the isolation of cDNA sequences
CC encoding human pancreatic tumour proteins. The polynucleotide
CC sequences encoding human pancreatic tumour proteins are useful for
CC stimulating an immune response in a patient and treating pancreatic
CC cancer in a patient. A host cell that expresses these polynucleotides
CC is useful for determining the presence of cancer in a patient. A
CC composition comprising the polynucleotide, its encoded protein, or an
CC antibody that binds to the protein may be used in the diagnosis,
CC prevention and/or treatment of diseases, particularly pancreatic
CC cancer. The sequences of the invention are also useful in pharmaceutical
CC compositions, e.g. vaccines, for the diagnosis and treatment of
CC pancreatic cancer. Such compositions may be useful for inhibiting the
CC development of cancer in a patient, or as markers for the progression
CC of cancer. The polynucleotide sequences may also be used as probes
CC or primers for nucleic acid hybridisation assays. AKK44061.ABK44209
CC represent cDNA sequences encoding for human pancreatic tumour proteins.
XX

Sequence 549 BP; 122 A; 167 C; 137 G; 119 T; 4 other;

Query Match	37.5%	Score 30.4;	DB 24;	Length 549;
Best Local Similarity	60.5%;	Pred. No. 1.8;		
Matches 49; Conservative	0;	Mismatches 32;	Indels 0	

QY 1 GGACCTCAGACACACCCAGGCCCCAGAGAAGATGGGCCCTCCTGTGACACAGTTTAA 60
 ||| ||||| | ||| | |||| | ||| |||| | ||||
 Db 406 GGATCTCAGAGTATACCAAGAGCCCTGTGTGAACCTGGGCAAGCTGGTCCTTTCAGGCCCT 460

QY 61 CCACATATACAGGAGAATA 81
||| | | | | |
Db 466 CCAGGACCTCCTGNGCTATA 486

```

RESULT 8
AAH48067
ID AAH48067 standard; cDNA; 1317 BP.
XX
AC AAH48067;
XX
DE 19-SEP-2001 (first entry)
XX
DE Murine HSP47 interacting protein, #1, coding sequence.
XX
XX Murine; heat shock protein interacting protein; HSP47; ss.
XX
OS Mus sp.
XX
PN JP2001145493-A.
XX
PD 29-MAY-2001.
XX
PF 19-NOV-1999; 99JP-0330631.
XX
PR 19-NOV-1999; 99JP-0330631.
XX
PA (KAGA-) KAGAKU GIJUTSU SHINKO JIGYODAN.
XX
DR WPI: 2001-395263/42.
XX
P-PSDB: MAG64211.
XX
PT Using the two-hybrid screening method to prepare proteins which
PT interact with the heat shock protein HSP47
XX
PS Claim 21; Page 18-19; 26pp; Japanese.
XX
CC The present invention relates to a method for preparing a protein which
CC interacts with the heat shock protein HSP47. The method involves the
CC two-hybrid screening method using the HSP47 gene and a mammalian cDNA
CC library. The present sequence is the coding sequence for a murine HSP47
CC interacting protein which was used in the present method. The HSP47
CC interacting proteins are useful for the diagnosis and treatment of
CC diseases caused by an increase or decrease in activity of HSP47.
XX
SQ Sequence 1317 BP; 275 A; 422 C; 410 G; 210 T; 0 other;
XX
Query Match 37.5%; Score 30.4; DB 22; Length 1317;
Best Local Similarity 71.4%; Pred. No. 2.3;
Matches 40; Conservative 0; Mismatches 16; Indels 0; Gaps 0;
OY 1 GGACCTCCAGACACCCAGGCCCGCAGAGAAGTGGGCGCTCTGTGACACGAG 56
DB 706 GGACCAACCGGGCGACGCGGCTCCTGGAGAGATGGGCGCCCTGTCACACGAG 761

```

```

XX
PF 01-DEC-1993; 93US-0159784.
XX
PR 01-DEC-1993; 93US-0159784.
XX
PA (HARD ) HARVARD COLLEGE.
XX
PI Oh SP, Olsen BR;
XX
DR WPI: 1997-350247/32.
XX
P-PSDB: AAM26327.
XX
PT Nucleic acid encoding human alpha-1 collagen - for production of
PT recombinant alpha-1 collagen, for use in the treatment of cartilage
PT degeneration
XX
PS Claim 1; Column 23-30; 35pp; English.
XX
CC A cDNA clone (AAH4484) codes for a human novel type alpha-1 (XVIII)
CC collagen (AAM26327) that is expressed in multiple tissues, especially
CC liver, lung and kidney. It was isolated from a placental cDNA
CC library using a probe based on an unidentified collagenous protein
CC and a probe based on mouse alpha-1 cDNA clone mc19. A claimed
CC plasmid comprising alpha-1 collagen nucleic acid and an expression
CC control sequence can be used to express recombinant collagen in
CC prokaryotic or eukaryotic (especially mammalian) host cells. The
CC collagen may be used to treat a patient suffering from a disease
CC associated with degradation of cartilage, and for supplementing
CC collagen.
XX
SQ Sequence 3394 BP; 668 A; 1130 C; 1080 G; 516 T; 0 other;
XX
Query Match 37.5%; Score 30.4; DB 18; Length 3394;
Best Local Similarity 71.4%; Pred. No. 2.9;
Matches 40; Conservative 0; Mismatches 16; Indels 0; Gaps 0;
OY 1 GGACCTCCAGACACCCAGGCCCGCAGAGAAGTGGGCGCTCTGTGACACGAG 56
DB 670 GGCCCTCCAGACCCCAAGGCCCGCAGAGAAGTGGGCGCCCGGACACACGAG 725

```

```

RESULT 9
AAT84484
ID AAT84484 standard; cDNA; 3394 BP.
XX
AC AAT84484;
XX
DE 19-NOV-1997 (first entry)
XX
DE Human alpha-1 collagen cDNA.
XX
XX Alpha-1 collagen; type XVIII collagen; cartilage degeneration; ss.
XX
OS Homo sapiens.
XX
PN Key 1.2055
XX CDS Location/Qualifiers
XX FT 1.2055
XX FT /*tag= a
XX PN US5643783-A.
XX PD 01-JUL-1997.

```

```

RESULT 10
AAH78379
ID AAH78379 standard; cDNA; 3394 BP.
XX
AC AAH78379;
XX
DE 25-AUG-1999 (first entry)
XX
DE Human alpha1 (XVIII) collagen cDNA.
XX
XX Alpha1(XVIII) collagen; mimetic; endostatin; atomic coordinate; library;
XX anti-angiogenic; heparin binding domain; receptor binding domain; mimetic;
XX alpha-helix A domain; carbohydrate recognition domain; CRD domain;
XX treatment; angiogenesis; tumour; human; ss.
XX
OS Homo sapiens.
XX
PN Key 1.2055
XX CDS Location/Qualifiers
XX FT 1.2055
XX FT /*tag= a
XX FT /product= "alpha1(XVIII) collagen"
XX FT /note= "Partial sequence, no start codon given"
XX PN WO931616-A1.
XX PD 24-JUN-1999.
XX PF 16-DEC-1998; 98WO-US26783.
XX PR 16-DEC-1997; 97US-0069727.
XX PA (HARD ) HARVARD COLLEGE.

```

XX HHenester E, Olsen BR, Sasaki T, Timpl R;
XX WPI: 1999-395243/33.
DR P-PSDB; AAY25113.
XX Identifying mimetics of mammalian endostatin
XX Disclosure; Fig 5A-C; 75pp; English.
PS
XX This invention describes a novel method for identifying mimetics of
CC mammalian endostatin. The method comprises identifying a compound
CC having atomic coordinates with non-trivial similarity to selected
CC coordinates of atoms of a mammalian endostatin involves (a) providing
CC a library of atomic coordinates of compounds in a library of candidate
CC compounds, (b) comparing the library of atomic coordinates to the
CC selected coordinates of a mammalian endostatin and (c) selecting from the
CC library at least one candidate compound on the basis of selection
CC criteria which include similarities between the atomic coordinates of the
CC selected candidate compound and the atomic coordinates of the mammalian
CC endostatin. The invention also describes the use of an anti-angiogenic
CC fragment of endostatin comprising a domain selected from a heparin
CC binding domain, a receptor binding domain, and exposed on alpha-helix A
CC domain, and a carbohydrate recognition domain (CRD) domain. The methods
CC can be used for designing and selecting endostatin mimics. The compounds
CC identified can be used for treating undesired angiogenesis, e.g. tumors.
CC This sequence encodes human alpha1(XVIII) collagen which is used in the
CC description of the method.
XX
SQ Sequence 3394 BP; 668 A; 1130 C; 1080 G; 516 T; 0 other;

Query Match 37.5%; Score 30.4; DB 20; Length 3394;
Best Local Similarity 71.4%; Pred. No. 2.9;
Matches 40; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

OY 1 GGACCTCCAGCACACCCCGCCAGAGAGTGCGGCTCCTGTGCACGAG 56
DB 670 GGCCCTCCAGAGCCCAAGGCGCCCAAGAGAGAGTGCGGCGCCCGCCGACACGAG 725

RESULT 11
ABN95680
ID ABN95680 standard; DNA: 3394 BP.
XX
AC ABN95680;
XX
DT 13-AUG-2002 (first entry)
XX
DE Gene #2178 used to diagnose liver cancer.
XX
KW Gene; liver cancer; ds; hepatocellular carcinoma; hepatotropic;
KW metastatic liver tumour; cytostatic; expression profile; disease state;
KW disease progression; drug toxicity; drug efficacy; drug metabolism.
XX
OS Homo sapiens.
XX
PN WO200229103-A2.
XX
PD 11-APR-2002.
XX
PF 02-OCT-2001; 2001MO-US30589.
XX
PR 02-OCT-2000; 2000US-237054P.
XX
PA (GENE-) GENE LOGIC INC.
XX
PI Horne D, Alvares C, Peres-Da-Silva S, Vockley JG;
XX WPI: 2002-426119/45.
DR
XX
PT Diagnosing and detecting the progression of liver cancer,
PT hepatocellular carcinoma or metastatic liver tumor in a patient,
PT involves detecting the level of expression of two or more genes in a

PT liver tissue sample -
XX
PS Claim 1; SEQ ID NO 2178; 298pp; English.
XX
CC The invention relates to a novel method for diagnosing and detecting the
CC progression of liver cancer, hepatocellular carcinoma or metastatic liver
CC tumour in a patient, and differentiating metastatic liver cancer from
CC hepatocellular carcinoma in a patient, involving detecting the level of
CC expression of two or more genes represented in ABN93503-ABN97455 in a
CC tissue sample. The method of the invention has hepatotropic, and
CC cytostatic activity. The method is useful for diagnosing and detecting
CC the progression of liver cancer, hepatocellular carcinoma and metastatic
CC liver carcinoma in a patient. The method is useful for identifying
CC expression profiles which serve as useful diagnostic markers as well as
CC markers that can be used to monitor disease states, disease progression,
CC drug toxicity, drug efficacy and drug metabolism.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at http://wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 3394 BP; 668 A; 1130 C; 1080 G; 516 T; 0 other;

Query Match 37.5%; Score 30.4; DB 24; Length 3394;
Best Local Similarity 71.4%; Pred. No. 2.9;
Matches 40; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

OY 1 GGACCTCCAGCACACCCCGCCAGAGAGTGCGGCTCCTGTGCACGAG 56
DB 670 GGCCCTCCAGAGCCCAAGGCGCCCAAGAGAGAGTGCGGCGCCCGCCGACACGAG 725

RESULT 12
AAH98343
ID AAH98343 standard; cDNA; 5467 BP.
XX
AC AAH98343;
XX
DT 12-OCT-2001 (first entry)
XX
DE Human EST-derived coding sequence SEQ ID NO: 200.
XX
XX Human; sheep; pig; cow; fruit fly; yeast; hamster; macaque; horse;
KW tomato; monkey; dog; sea urchin; expressed sequence tag; EST;
KW diagnostics; forensic test; gene mapping; genetic disorder;
KW biodiversity; gene therapy; nutrition; ss.
XX
OS Homo sapiens.
XX
PN WO200154477-A2.
XX
PD 02-AUG-2001.
XX
PF 25-JAN-2001; 2001MO-US02687.
XX
PR 25-JAN-2000; 2000US-0491404.
PR 17-JUL-2000; 2000US-0617746.
PR 03-AUG-2000; 2000US-0631451.
PR 15-SEP-2000; 2000US-0663870.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Zhou P, Qian XB, Wang Z, Chen R, Asundi V;
PI Cao Y, Drmanac RA, Zhang J, Werhman T;
XX WPI: 2001-476164/51.
DR P-PSDB; AAM23684.
XX
XX Isolated polypeptide for treatment of diseases, diagnostics, raising
PT antibodies and research use -
XX
PS Claim 1; Page 314-315; 1275pp; English.
XX
PT The present invention provides the protein and coding sequences of novel

CC proteins from a variety of organisms, including human, dog, cat, horse,
CC cow, pig, hamster, monkey, macaque, yeast, bacteria, fruit fly, sea
CC urchin and tomato. These were derived from expressed sequence tags (ESTs)
CC from the organism of interest. They can be used in diagnostics,
CC forensics, gene mapping, identification of mutations, to assess
CC biodiversity and for nutritional purposes. The present sequence is a cDNA
CC of the invention.

XX
SQ Sequence 5467 BP; 1330 A; 1323 C; 1550 G; 1264 T; 0 other;

Query Match 37.3%; Score 30.2; DB 22; Length 5467;
Best Local Similarity 62.7%; Pred. No. 3.8;
Matches 47; Conservative 0; Mismatches 28; Indels 0; Gaps 0;

OY 1 GGACCTCCAGCACACCCAGGCCCCAGAGAACTGGGCTCTGTCACACAGTTTA 60
DB 634 GGCCCCCAGGCCGCCGCCGCTGTACATCTGTCATCTCTGTTCCCGATCT 693

OY 61 CCACAATATACAGGA 75

DB 694 CCAGGATACCAAGGA 708

RESULT 13

AAH98411
ID AAH98411 standard; cDNA; 5468 BP.

XX
AC AAH98411;

XX
DT 12-OCT-2001 (first entry)

DE Human EST-derived coding sequence SEQ ID NO: 268.

XX Human; sheep; pig; cow; fruit fly; yeast; hamster; macaque; horse;
KW tomato; monkey; dog; sea urchin; expressed sequence tag; EST;
KW diagnostics; forensic test; gene mapping; genetic disorder;
KW biodiversity; gene therapy; nutrition; ss.

OS Homo sapiens.

XX
PN WO200154477-A2.

XX
PD 02-AUG-2001.

XX
PF 25-JAN-2001; 2001WO-US02687.

XX
PR 25-JAN-2000; 2000US-0491404.

XX
PR 17-JUL-2000; 2000US-0617746.

XX
PR 03-AUG-2000; 2000US-0631451.

XX
PR 15-SEP-2000; 2000US-0663870.

XX
PA (HYSE-) HYSEQ INC.

XX
PI Tang YT, Liu C, Zhou P, Qian XB, Wang Z, Chen R, Asundi V;

XX
PI Cao Y, Drmanac RA, Zhang J, Werhman T;

XX
PI WPI; 2001-476164/51.

XX
PI P-PSDB; AAM23752.

XX
PT Isolated polypeptide for treatment of diseases, diagnostics, raising

XX
PT antibodies and research use -

XX
PS Claim 1; Page 381-382; 1275pp; English.

XX
CC The present invention provides the protein and coding sequences of novel
CC proteins from a variety of organisms, including human, dog, cat, horse,
CC cow, pig, hamster, monkey, macaque, yeast, bacteria, fruit fly, sea
CC urchin and tomato. These were derived from expressed sequence tags (ESTs)
CC from the organism of interest. They can be used in diagnostics,
CC forensics, gene mapping, identification of mutations, to assess
CC biodiversity and for nutritional purposes. The present sequence is a cDNA
CC of the invention.

SQ Sequence 5468 BP; 1330 A; 1323 C; 1551 G; 1264 T; 0 other;

Query Match 37.3%; Score 30.2; DB 22; Length 5468;
Best Local Similarity 62.7%; Pred. No. 3.8;
Matches 47; Conservative 0; Mismatches 28; Indels 0; Gaps 0;

OY 1 GGACCTCCAGCACACCCAGGCCCCAGAGAACTGGGCTCTGTCACACAGTTTA 60

DB 634 GGCCCCCAGGCCGCCGCCGCTGTACATCTGTCATCTCTGTTCCCGATCT 693

OY 61 CCACAATATACAGGA 75

DB 694 CCAGGATACCAAGGA 708

RESULT 14

AAS44890
ID AAS44890 standard; DNA; 707 BP.

XX
AC AAS44890;

XX
DT 18-DEC-2001 (first entry)

DE Human contig polynucleotide sequence #143.

XX Mammal; human; rhesus monkey; baker's yeast; fission yeast; Norway rat;
KW mouse; Chinese hamster; African clawed frog; fruit fly; dog; leukemia;
KW cancer; lymphoma; neuroblastoma; autoimmune disorder; cell proliferation;
KW nervous system disorder; inflammatory disorder; cell differentiation; ds;
KW angiogenesis; stem cell growth factor; activin; inhibin; cartilage; burn;

KW genetic disorder; bone regeneration; tendon; ligament; tissue repair;

KW cytoskeletal; antirheumatic; antiarthritic; vulnery; antiinflammatory;

KW antibacterial; immunosuppressive; vasotropic; antiparkinsonian;

KW neuroprotective; osteopathic; antidiabetic; antistimatic; antiallergic;

XX
OS Homo sapiens.

XX
PN WO200164834-A2.

XX
PD 07-SEP-2001.

XX
PF 26-FEB-2001; 2001WO-US04926.

XX
PR 28-FEB-2000; 2000US-0515126.

XX
PR 18-MAY-2000; 2000US-0577409.

XX
PR 17-JUN-2000; 2000US-0597707.

XX
PR 14-JUL-2000; 2000US-0616807.

XX
PR 19-SEP-2000; 2000US-0664641.

XX
PA (HYSE-) HYSEQ INC.

XX
PI Tang YT, Liu C, Zhou P, Asundi V, Zhang J, Zhao QA, Ren F;

XX
PI Xue AJ, Yang Y, Wehrman T, Wang J, Ma Y, Wang D, Chen R, Xu C;

XX
PI Drmanac R;

XX
PI WPI; 2001-589862/66.

XX
PI P-PSDB; AAU27990.

XX
PT Novel polypeptides and nucleic acids obtained from cDNA libraries

XX
PT prepared from various human tissues, for diagnosis, treatment of

XX
PT cancer, neurological, inflammatory disorders and for use in arrays for

XX
PT detection -

XX
PS Claim 1; SEQ ID NO 487; 153pp; English.

XX
CC Sequences AAS44576-AAS44919 represent full-length polynucleotides and

CC contig polynucleotides encoding polypeptides of the invention. The DNA

CC and protein sequences are useful for the treatment, diagnosis and

CC prevention of various types of disorder in a mammalian subject such as a

CC human, dog, monkey, mouse, hamster or rat. The disorders include cancers

CC such as leukemia, lymphoma and neuroblastoma, autoimmune disorders such

PR	07-JUL-2000	2000US-02116880
PR	11-JUL-2000	2000US-02117687
PR	11-JUL-2000	2000US-02117696
PR	14-JUL-2000	2000US-02118290
PR	26-JUL-2000	2000US-02209693
PR	26-JUL-2000	2000US-02209694
PR	1-AUG-2000	2000US-02245119
PR	1-AUG-2000	2000US-02245139
PR	1-AUG-2000	2000US-02255114
PR	1-AUG-2000	2000US-02255114
PR	1-AUG-2000	2000US-02255166
PR	1-AUG-2000	2000US-02255677
PR	1-AUG-2000	2000US-02255688
PR	1-AUG-2000	2000US-02255710
PR	1-AUG-2000	2000US-02255747
PR	1-AUG-2000	2000US-02255758
PR	1-AUG-2000	2000US-02257159
PR	1-AUG-2000	2000US-02257558

PR	22-AUG-2000;	2000US-0226681.
PR	22-AUG-2000;	2000US-0226868.
PR	22-AUG-2000;	2000US-0227182.

PR	30-AUG-2000;	2000US-0228924.
PR	01-SEP-2000;	2000US-0229287.
PR	01-SEP-2000;	2000US-0229343.

PR	05-SEP-2000;	2000US-0229513
PR	05-SEP-2000;	2000US-0229509
PR	01-SEP-2000;	2000US-0229345
PR	01-SEP-2000;	2000US-0229347

PR	08-SEP-2000;	2000US-0231242.
PR	08-SEP-2000;	2000US-0231243.
PR	08-SEP-2000;	2000US-0231244.

PR	08-SEP-2000;	2000US-0231414.
PR	08-SEP-2000;	2000US-0232080.
PR	08-SEP-2000;	2000US-0232081.

PR 14-SEP-2000; 2000US-0232397.

PR 14-SEP-2000; 2000US-02332399
PR 14-SEP-2000; 2000US-02332401
PR 14-SEP-2000; 2000US-02332410
PR 14-SEP-2000; 2000US-02333064
PR 14-SEP-2000; 2000US-02333663
PR 14-SEP-2000; 2000US-02333665
PR 14-SEP-2000; 2000US-02342223
PR 21-SEP-2000; 2000US-02342274
PR 21-SEP-2000; 2000US-02343474
PR 25-SEP-2000; 2000US-02343997
PR 25-SEP-2000; 2000US-02349998
PR 25-SEP-2000; 2000US-02349998

PR 26-SEP-2000; 2000US-0235484;
PR 27-SEP-2000; 2000US-0235834;

PR	27-SEP-2000; 2000US-02330636.
PR	29-SEP-2000; 2000US-0236327.

PR	29-SEP-2000; 2000US-0236368.
PR	23 SEP 2000; 2000US-0230307.

PR 29-SEP-2000; 2000US-0236370.

PR 02-OCT-2000; 2000US-0237037.

PR 02-OCT-2000; 2000US-0237039.

PR 13-OCT-2000; 2000US-0239935.

PR 20-OCT-2000; 2000US-0240960.

PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786

PR	20-OCT-2000; 2000US-0241787.
PR	20-OCT-2000; 2000US-0241809

PR 20-OCT-2000; 2000US-0241809.
 PR 20-OCT-2000; 2000US-0241826.
 PR 01-NOV-2000; 2000US-0244617.
 PR 08-NOV-2000; 2000US-0246474.
 PR 08-NOV-2000; 2000US-0246475.
 PR 08-NOV-2000; 2000US-0246476.
 PR 08-NOV-2000; 2000US-0246477.
 PR 08-NOV-2000; 2000US-0246478.
 PR 08-NOV-2000; 2000US-0246523.
 PR 08-NOV-2000; 2000US-0246524.
 PR 08-NOV-2000; 2000US-0246525.
 PR 08-NOV-2000; 2000US-0246526.
 PR 08-NOV-2000; 2000US-0246527.
 PR 08-NOV-2000; 2000US-0246528.
 PR 08-NOV-2000; 2000US-0246532.
 PR 08-NOV-2000; 2000US-0246609.
 PR 08-NOV-2000; 2000US-0246610.
 PR 08-NOV-2000; 2000US-0246611.
 PR 08-NOV-2000; 2000US-0246613.
 PR 17-NOV-2000; 2000US-0249207.
 PR 17-NOV-2000; 2000US-0249208.
 PR 17-NOV-2000; 2000US-0249209.
 PR 17-NOV-2000; 2000US-0249210.
 PR 17-NOV-2000; 2000US-0249211.
 PR 17-NOV-2000; 2000US-0249212.
 PR 17-NOV-2000; 2000US-0249213.
 PR 17-NOV-2000; 2000US-0249214.
 PR 17-NOV-2000; 2000US-0249215.
 PR 17-NOV-2000; 2000US-0249216.
 PR 17-NOV-2000; 2000US-0249217.
 PR 17-NOV-2000; 2000US-0249218.
 PR 17-NOV-2000; 2000US-0249219.
 PR 17-NOV-2000; 2000US-0249220.
 PR 17-NOV-2000; 2000US-0249221.
 PR 17-NOV-2000; 2000US-0249222.
 PR 17-NOV-2000; 2000US-0249223.
 PR 17-NOV-2000; 2000US-0249224.
 PR 17-NOV-2000; 2000US-0249225.
 PR 17-NOV-2000; 2000US-0249226.
 PR 17-NOV-2000; 2000US-0249227.
 PR 17-NOV-2000; 2000US-0249228.
 PR 17-NOV-2000; 2000US-0249229.
 PR 17-NOV-2000; 2000US-0249230.
 PR 17-NOV-2000; 2000US-0250160.
 PR 01-DEC-2000; 2000US-0250391.
 PR 05-DEC-2000; 2000US-0251030.
 PR 05-DEC-2000; 2000US-0251988.
 PR 05-DEC-2000; 2000US-0256719.
 PR 06-DEC-2000; 2000US-0251479.
 PR 08-DEC-2000; 2000US-0251856.
 PR 08-DEC-2000; 2000US-0251869.
 PR 08-DEC-2000; 2000US-0251869.
 PR 08-DEC-2000; 2000US-0251989.
 PR 08-DEC-2000; 2000US-0251990.
 PR 11-DEC-2000; 2000US-0254097.
 PR 05-JAN-2001; 2001US-0259678.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Rosen CA, Barash SC, Ruben SM;
 XX
 DR WPI: 2001-465460/50.
 DR P-PSDB; ANU17096.
 XX
 PT Novel polypeptides useful for diagnosing, treating, preventing and/or
 PT prognosing disorders related to the proteins, including cancers, immune
 PT disorders and neuronal disorders
 XX
 PS Claim 1: SEQ ID No 48; 880pp; English.
 XX
 CC The invention relates to novel isolated polypeptides (I), and
 CC polynucleotides (II). (I), (II) and the antibody to (I) are useful for
 CC diagnosing, preventing and treating diseases including immune system
 CC disorders (e.g. congenital and acquired immunodeficiencies, autoimmune
 CC disorders (e.g. rheumatoid arthritis), inflammatory conditions, organ
 CC transplant rejections and graft versus host disease, infectious diseases
 CC (e.g. hepatitis C), bleeding disorders, haemoglobin abnormalities and
 CC other blood-related disorders (sickle cell anaemia), myeloproliferative

CC disorders, primary haematopoietic disorders, hyperproliferative
 CC disorders (e.g. Gaucher's disease and cancer), neurodegenerative
 CC disorders (e.g. Alzheimer's disease, Parkinson's disease), chromosomal
 CC abnormalities (Down syndrome), ischaemic injury (e.g. stroke), renal
 CC disorders (e.g. glomerulonephritis), cardiovascular disorders
 CC (e.g. arrhythmia), respiratory disorders, dermatological disorders,
 CC wound healing, epithelial cell proliferation, endocrine disorders (e.g.
 CC Addison's disease), reproductive system disorders, gastrointestinal
 CC disorder (inflammatory disorders), liver disorders (cirrhosis),
 CC as stimulators of B-cell responsiveness to pathogens, activators of
 CC T-cells, to induce higher affinity antibodies, and as a means to induce
 CC tumour proliferation in pathologies e.g. acquired immune deficiency
 CC syndrome (AIDS). AAS26976-AAS27850 represent novel signal transduction
 CC pathway protein coding sequences and PCR primers of the invention.
 XX

Query Match 37.0%; Score 30; DB 22; Length 2686;
 Best Local Similarity 61.5%; Pred. No. 3.7;
 Matches 48; Conservative 0; Mismatches 30; Indels 0; Gaps 0;

QY 1 GGACCTCCAGCACACCCAGGCCCCAGAGAAGTGGGCTCTCTGTGACACAGTTTA 60
 db 1305 GGAGCTCCTGCACACCGAGAGACCTATGTGAAGCGGCTGCACCTCTGGACACGTTT 1364
 QY 61 CCACATATACAGAGAA 78
 db 1365 CTGCACAGGCTGACCGA 1382

Search completed: February 19, 2003, 22:56:13
 Job time : 65.0336 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 19, 2003, 23:16:26 ; Search time 258.251 Seconds

(without alignments)
9128.055 Million cell updates/sec

Title: US-09-997-610-1_COPY_56_136

Perfect score: 81

Sequence: 1 ggaccctccagcacacccacg.....cacatatacagagagaata 81

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapept 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: gb_ba:**

2: gb_hgt:**

3: gb_in:**

4: gb_om:**

5: gb_ov:**

6: gb_pat:**

7: gb_ph:**

8: gb_pl:**

9: gb_pr:**

10: gb_ro:**

11: gb_sts:**

12: gb_sy:**

13: gb_un:**

14: gb_vl:**

15: em_ba:**

16: em_fun:**

17: em_hum:**

18: em_in:**

19: em_mu:**

20: em_om:**

Result No.	Score	Query Match	Length	DB	ID	Description
1	57	70.4	145880	9	HS302D9	282198 Human DNA s
2	35.8	44.2	287	3	008420010	AJ420010 Osterlagl
3	35	43.2	694	3	HSNCO11	X61045 Hydra N-COL
4	34.8	43.0	1993	3	BMCOLGMR	Z30348 B.mori mRNA
5	33.4	41.2	3734	14	HVSX99519	X99519 Herpesvirus
6	33.4	41.2	6522	3	US8736	U58736 Caenorhabd
7	33.2	41.0	36532	3	CEP5781	Z78064 Caenorhabd
8	33.2	41.0	269619	3	CEY5184A	AL132952 Caenorhab
9	33	40.7	5851	3	AF282902	AF282902 Hydra vul
10	32.6	40.2	1776	10	RNU57362	U57362 Rattus norv
11	32	39.5	3612	10	MMCOL4A	X06777 Mouse mRNA
12	32	39.5	6512	6	AX306181	AX306181 Sequence
13	32	39.5	6512	10	MUSCOL1A4A	J04694 Mus musculu
14	32	39.5	15143	9	HS404862	AJ004862 Homo sapi
15	32	39.5	100516	2	HS5171M_3	Continuation (4 of
16	32	39.5	340000	9	HS21C102	AL163302 Homo sapi
17	31.8	39.3	1405	5	AF378825	AF378825 Danio rer
18	31.8	39.3	1481	5	AF378826	AF378826 Danio rer
19	31.8	39.3	2814	5	AF378824	AF378824 Danio rer
20	31.8	39.3	3846	5	AF349034	AF349034 Danio rer
21	31.8	39.3	6652	14	HSV488A	M55264 Herpesvirus
22	31.8	39.3	105936	3	AC084440	AC084440 Caenorhab
23	31.6	39.0	1798	9	HUMNACITL	M63597 Human fibril
24	31.6	39.0	2157	9	HUMNCO1L	L12347 Homo sapien
25	31.6	39.0	2247	3	CBU84501	U84501 Caenorhabd
26	31.6	39.0	3720	3	HUMALIXXC	D38163 Human mRNA
27	31.6	39.0	34397	3	AF000198	AF000198 Caenorhab
28	31.6	39.0	43559	3	CBRG47L22	AC084677 Caenorhab
29	31.4	38.8	862	10	RNU57361	U57361 Rattus norv
30	31.2	38.5	508	9	HUMCOL1A16	M26553 Human alpha
31	31.2	38.5	2853	9	HSCOL4A1	X05561 Human mRNA
32	31.2	38.5	3636	9	HSC4A1	Y00706 Human mRNA
33	31.2	38.5	5010	6	AX393321	AX393321 Sequence
34	31.2	38.5	31110	3	CBRG14K24	AC084504 Caenorhab
35	31.2	38.5	160149	2	AC016380	AC016380 Homo sapi
36	31.2	38.5	166266	2	AP001146	AP001146 Homo sapi
37	31.2	38.5	185404	2	AC025409	AC025409 Homo sapi
38	31.2	38.5	186120	9	ALJ90755	ALJ90755 Human DNA
39	31	38.3	9448	3	CEC022327	U22327 Caenorhabd
40	31	38.3	9541	3	CEC0LA2IV	U22327 Caenorhabd
41	31	38.3	34671	3	U53342	U53342 C.elegans a
42	31	38.3	87849	3	AC084453	AC084453 Caenorhab
43	30.8	38.0	1395	5	DRE318214	AJ318214 Danio rer
44	30.8	38.0	42404	3	CBRG46J06	AC084656 Caenorhab
45	30.4	37.5	2690	10	AB085837	AB085837 Mus muscu

ALIGNMENTS

RESULT 1	LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM	REFERENCE	AUTHORS	TITLE
HS302D9	HS302D9	Human DNA sequence from clone RPI-302D9 on chromosome 22 Contains	282198	282198.2	GI:6572207	HTG.	Homo sapiens.	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.	Bridgeman, A.	Direct Submission

JOURNAL

COMMENT

Submitted (08-DEC-1999) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humqueyes@sanger.ac.uk
Requests: clonerequest@sanger.ac.uk
On Dec 13, 1999 this sequence version replaced g1:3164067.
During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.
This sequence has been finished according to sequence map criteria as follows. An attempt is made to resolve all sequencing problems, such as compressions and repeats, but not necessarily within known annotated human repeat sequence elements (e.g. Alu). Where the sequence is ambiguous, there is an annotation using the 'unsure' feature key.
The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em: EMBL, Sw: SWISSPROT, Tr: TREMBL, Wp: WORMPEP; information on the WORMPEP database can be found at <http://www.sanger.ac.uk/projects/C-elegans/wormpep> This sequence was generated from part of bacterial clone contigs of human chromosome 22, constructed by the Sanger Centre Chromosome 22 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/Chr22>
Rpl-302D9 is from the library RPl-1 constructed at the Roswell Park Cancer Institute by the group of Pieter de Jong. For further details see <http://bacpac.med.buffalo.edu/VECTOR:pcypac2>
This sequence is the entire insert of clone RPl-302D9 The true left end of clone CNA-282F2 is at 69682 in this sequence. The true right end of clone CNA-415G2 is at 55167 in this sequence.

FEATURES

source
1. .145880
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="22"
/clone="RPl-302D9"
/clone_lib="RPl-1"
188. .245
/note="MER3 repeat: matches 144. .209 of consensus"
246. .571
/note="AluX repeat: matches 1. .312 of consensus"
572. .759
/note="MER3 repeat: matches 1. .144 of consensus"
783. .933
/note="MER5A repeat: matches 26. .187 of consensus"
1033. .1336
/note="AluSp repeat: matches 1. .299 of consensus"
1450. .1583
/note="MIR repeat: matches 24. .160 of consensus"
1687. .1752
/note="L2 repeat: matches 2593. .2661 of consensus"
2350. .2660
/note="AluSc repeat: matches 3. .309 of consensus"
2684. .2981
/note="AluSq repeat: matches 2. .300 of consensus"
3323. .3343
/note="MLTIE repeat: matches 116. .136 of consensus"
3344. .3652
/note="AluY repeat: matches 1. .309 of consensus"
3653. .3928
/note="MLTIE repeat: matches 136. .359 of consensus"
3929. .4278
/note="THE1B repeat: matches 3. .364 of consensus"
4279. .4485
/note="MLTIE repeat: matches 359. .568 of consensus"
5073. .5176
/note="52 copies 2 mer ct 78 conserved"
5181. .5491
/note="AluJb repeat: matches 1. .311 of consensus"
6369. .6485
/note="L2 repeat: matches 2579. .2705 of consensus"

repeat_region 6647. .6685
/note="MADE1 repeat: matches 1. .23 of consensus"
6686. .6987
/note="AluX repeat: matches 1. .302 of consensus"
6988. .7036
/note="MADE1 repeat: matches 23. .77 of consensus"
7482. .7754
/note="AluJb repeat: matches 9. .290 of consensus"
7775. .8060
/note="AluJo repeat: matches 1. .295 of consensus"
8414. .8551
/note="L2 repeat: matches 2553. .2706 of consensus"
8914. .9030
/note="MIR repeat: matches 147. .262 of consensus"
9110. .9280
/note="MIR repeat: matches 91. .262 of consensus"
9283. .9412
/note="MIR repeat: matches 15. .144 of consensus"
9521. .9679
/note="FAM repeat: matches 3. .161 of consensus"
9820. .10225
/note="MSTR repeat: matches 2. .425 of consensus"
/note="match: GSS: Em:856592"
/note="match: GSS: Em:10728"
/note="match: GSS: Em:AQ701486"
/note="match: GSS: Em:AQ249. .10706"
/note="match: GSS: Em:AQ225495"
10312. .10383
/note="MIR repeat: matches 79. .150 of consensus"
10718. .11310
/note="match: GSS: Em:BI4024"
10784. .11201
/note="match: GSS: Em:B43656"
11838. .11946
/note="MIR repeat: matches 20. .137 of consensus"
12174. .12445
/note="L2 repeat: matches 1988. .2275 of consensus"
12444. .12642
/note="MIR repeat: matches 63. .241 of consensus"
13017. .13369
/note="match: STR: Em:G49301"
13331. .13397
/note="MIR repeat: matches 174. .244 of consensus"
13398. .13698
/note="AluSp repeat: matches 1. .302 of consensus"
13699. .13810
/note="MIR repeat: matches 76. .174 of consensus"
13806. .13919
/note="MIR repeat: matches 77. .189 of consensus"
13945. .14060
/note="MIR repeat: matches 24. .142 of consensus"
14061. .14367
/note="AluY repeat: matches 1. .301 of consensus"
14368. .14452
/note="MIR repeat: matches 141. .225 of consensus"
14589. .14679
/note="MIR repeat: matches 173. .262 of consensus"
14597. .15201
/note="match: GSS: Em:AQ553482"
14616. .15060
/note="match: GSS: Em:AQ370601"
14868. .15040
/note="MIR repeat: matches 49. .233 of consensus"
15071. .15188
/note="L2 repeat: matches 2112. .2239 of consensus"
15304. .15399
/note="MLTIB repeat: matches 1. .99 of consensus"
15490. .15662
/note="AluSq repeat: matches 2. .114 of consensus"
15669. .15727
/note="MLTIB repeat: matches 119. .178 of consensus"
15728. .16027

RESULT 7
CEE57BL/c 36532 bp DNA linear INV 23-JAN-2002
LOCUS Caenorhabditis elegans cosmid F57BL, complete sequence.
DEFINITION 278064
ACCESSION 278064.1 GI:1487912
VERSION HTG; Collagen.
KEYWORDS Caenorhabditis elegans.
SOURCE Caenorhabditis elegans.
ORGANISM Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
Rhabditidae; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE 1
AUTHORS none.
TITLE Genome sequence of the nematode C. elegans: a platform for
investigating biology. The C. elegans Sequencing Consortium
JOURNAL Science 282 (5396), 2012-2018 (1998)
MEDLINE 99069613
PUBMED 9851916
REMARK The C.elegans Sequencing Consortium.
REFERENCE 2 (bases 1 to 36532)
AUTHORS Sims,M.A.
TITLE Direct Submission
JOURNAL Submitted (09-AUG-1996) Nematode Sequencing Project, Sanger
Institute, Hinxton, Cambridge CB10 1SA, England and Department of
Genetics, Washington University, St. Louis, MO 63110, USA. E-mail:
jes@anger.ac.uk or tw@nematode.wustl.edu
Coding sequences below are predicted from computer analysis, using
predictions from GeneFinder (P. Green, U. Washington), and other
available information.
Current sequence finishing criteria for the C. elegans genome
sequencing consortium are that all bases are either sequenced
unambiguously on both strands, or on a single strand with both a
dye primer and dye terminator reaction, from distinct subclones.
Exceptions are indicated by an explicit note.
IMPORTANT: This sequence is not the entire insert of clone F57BL.
It may be shorter because we only sequence overlapping sections
once, or longer because we arrange for a small overlap between
neighbouring submissions.
The true left end of clone F57BL is at 1 in this sequence. The true
right end of clone F57BL is at 5273 in
sequence 278060.
The true left end of clone C34D1 is at 35789 in this sequence. The
start of this sequence (1..104) overlaps with the end of sequence
278065.
The end of this sequence (35789..36532) overlaps with the start of
sequence 278060.
For a graphical representation of this sequence and its analysis
see: [http://wormbase.sanger.ac.uk/perl/ace/elegans/seq/sequence?](http://wormbase.sanger.ac.uk/perl/ace/elegans/seq/sequence?name=F57BL)
name=F57BL
IMPORTANT: This sequence is NOT necessarily the entire insert of
the specified clone. It may be shorter because we only sequence
overlapping sections once, or longer because we arrange for a small
overlap between neighbouring submissions.
location/Qualifiers
1..36532
/organism="Caenorhabditis elegans"
/db_xref="taxon:6239"
/chromosome="V"
/clone="F57BL"
complement(join(3851..4162,4210..4629,4674..4877,
4921..5024,5071..5452))
/gene="F57BL.2"
complement(join(3851..4162,4210..4629,4674..4877,
4921..5024,5071..5452))
/gene="F57BL.2"
/note="predicted using GeneFinder
cDNA EST Yk66999.3 comes from this gene
cDNA EST Yk66999.5 comes from this gene"
/codon_start=1
/protein_id="CAB01511.1"
/db_xref="GI:3877753"
/db_xref="SPTREMBL:Q20924"

/translation="MALRHTISPFQFSNRHSPPTRTSYRTGVQPLDSTPYVRDSQ
PGITTTGTQRFHESADOSEITOLNSKFTVEHEFSYKETSMTKEMWOWLETRIMWR
RRVPTWAQFRKRLMAVYLFAMLYKARAOCLPQGTIHNSGSAVDANNASEKQKH
QTTISNLRAEFSADKOLDEKTDKLELVNLEHSSKMSKSAIEELKQIKLMOAIED
ALOOMKEIDDAKSTKIHSTPEKAPATAPASIPSPSOQPHITRALLGVAVAS
LIGASTOHSSSRPVSATKDGEPFDMSPTREGVALLRDVLSREACTYDKAR
LTVKARLVTPKSVSTQHRWMSGIVPRHAKLLDVACTDSCCTKMDPLPVEYKRR
DGSYOEQEQFCSVPTIQNHSPINHVQFRFRNHGDMKTCAYLIRYGEVPOPKETQ
PMTDNGTESKLESIAVNSVETA"
complement(join(5920..6069,6126..6209,6280..6408,
6453..6566))
/gene="F57BL.1"
complement(join(5920..6069,6126..6209,6280..6408,
6453..6566))
/gene="F57BL.1"
/note="predicted using GeneFinder"
/codon_start=1
/protein_id="CAB01510.1"
/db_xref="GI:3877752"
/db_xref="SPTREMBL:Q20923"
/translation="MNNQCYEDERILMKHYVLSCTSSGIRNIAPPTIVOTIELILO
FILPILGIGYIAIIRIKILMKRAALNKYELTILOKAISSSVFLCOTIFETIATFL
IKRINTTEICAGAAPPCPEFFTSKEIRKLVSVSLTQGSANSQORQTSRAI"
complement(join(10251..10976,11025..11240))
/gene="F57BL.3"
complement(join(10251..10976,11025..11240))
/gene="F57BL.3"
/note="predicted using GeneFinder"
contains similarity to Pfam domain: PF01391 (collagen
triple helix repeat (20 copies)). Score=45.4,
E-value=4e-10, N=3; PF01484 (Nematode cuticle collagen
N-terminal domain), Score=40.6, E-value=1.1e-08, N=1
cDNA EST Yk73c7.5 comes from this gene"
/codon_start=1
/protein_id="CAB01509.1"
/db_xref="GI:3877751"
/db_xref="SPTREMBL:Q20922"
/translation="WSSEOPKQIAYETRESLRKVAFCIAVSTATLTATVAVPMLINY
MQHVOSLQSEVEFCASHRNSGLNDEYDFRFGSVGNARIRREYHRRARASKVRQSY
GADAAYGFCGSGAGSCSCGTGAAGACGAGDGAAPNGAGAGAPNPQDSQDT
AGDPFCFCPPAGPPGPGAGPGGKPSGAGVGGTGCASLPPPGPAGPPGSGQFG
SNGNAGAPGAPGGVGVPPGPPGAGPPGAGPPGPPGQSGSGQPGCPGQDGA
PGAPGPGAGAGCGQDSGSEBACHHCPPPRAPG"
complement(join(13023..13760,13809..13948,14082..14274))
/gene="F57BL.4"
complement(join(13023..13760,13809..13948,14082..14274))
/gene="F57BL.4"
/note="predicted using GeneFinder
contains similarity to Pfam domain: PF01391 (collagen
triple helix repeat (20 copies)). Score=44.2,
E-value=9.7e-10, N=3; PF01484 (Nematode cuticle collagen
N-terminal domain), Score=40.6, E-value=1.1e-08, N=1
cDNA EST Yk93c11.3 comes from this gene
cDNA EST Yk70b10.3 comes from this gene
cDNA EST Yk93c11.5 comes from this gene
cDNA EST Yk102e6.3 comes from this gene
cDNA EST Yk100f2.5 comes from this gene
cDNA EST Yk99e7.3 comes from this gene
cDNA EST Yk99e7.5 comes from this gene
cDNA EST Yk72e11.3 comes from this gene
cDNA EST Yk70b10.5 comes from this gene
cDNA EST Yk102e6.5 comes from this gene
cDNA EST Yk72e11.5 comes from this gene
cDNA EST Yk170f7.3 comes from this gene
cDNA EST Yk170f7.5 comes from this gene
cDNA EST Yk279e6.3 comes from this gene
cDNA EST Yk279e6.5 comes from this gene"
/codon_start=1
/protein_id="CAB01508.1"
/db_xref="GI:3877750"
/db_xref="SPTREMBL:Q20921"
/translation="MAEHGCVIDKAVEGCGKPSNPEFLASLPYVYSEVYTIASI
KGMTROMSNHLTRSCPFHSFTIATLTATVAVPMLINYMQHVOSLQSEVEFCARHS

NGLMDEYORFECVSGVAGRIKRESYHRRASGPHRASVSKYRQSGADAAVCGFGSAGC
 SCCSGCTGACGAGSPGODGAPNDGAPGAPNPDASSEDPTAGDPSCPCPAPGP
 GPGSAPGCGPGSAPGAGPGGSGASLPGPPGAPGPGPGSGGSGMNGAPGAPGVV
 DVGCPGAPGPGPGPAGAPGPGGSGASGAGCGGCGPGDACAAPGAPGPGGACAPGQ
 DDDSGSEACDHCPPPRTRAPG"
 join(19688..19768,19816..19867,19913..20010,20397..20489,
 20538..20687)
 /gene="F57B1.5"
 join(19688..19768,19816..19867,19913..20010,20397..20489,
 20538..20687)
 /gene="F57B1.5"
 /note="predicted using GeneFinder"
 /codon_start=1
 /protein_id="CAB01512.1"
 /db_xref="GI:3877754"
 /db_xref="SPTREMBL:Q20925"
 /translation="MPGKRFYKFPVVPFNVLSTSKITIIILLISLYGMDTKORIS
 GSNFLAIYDILAIIGIASISLMEIVQKTYKWEVSEFNIIGILLASISTSTITLD
 PASIOSLTHILILIMIFSAPWAMRETYNNLSISEWGCGEETLEILBOH"
 join(23062..23112,24099..24151,24322..24499,24547..24696,
 24748..24830,25045..25267)
 /gene="F57B1.6"
 join(23062..23112,24099..24151,24322..24499,24547..24696,
 24748..24830,25045..25267)
 /gene="F57B1.6"
 /note="predicted using GeneFinder"
 cDNA EST EMBL:T01341 comes from this gene"
 /codon_start=1
 /protein_id="CAB01513.1"
 /db_xref="GI:3877755"
 /db_xref="SPTREMBL:Q20926"
 /translation="MTMNYCGITKILNVDAVQCKLQSFSAVTLHLCEPLGDSRL
 YSRLLIDRKSEPIYKQHQIVSISGLCTSSVDFIYVSLLYTVDPKVLINPRM
 FPLWPSSTLILSLFLSAGPRATKFAVDNOMITSLIHALMGACILSSVS
 ALCQANDADIAVAFAFPAQFQLASAWYSRLRALVLSIILNGIIVSTYLVGI
 SKYRTFQLAKEPPQOGLIERTYFA"
 complement(join(34555..34692,34753..35192,35241..35370))
 /gene="F57B1.7"
 complement(join(34555..34692,34753..35192,35241..35370))
 /gene="F57B1.7"
 /note="predicted using GeneFinder"
 /codon_start=1
 /protein_id="CAB01507.1"
 /db_xref="GI:3877749"
 /db_xref="SPTREMBL:Q20920"
 /translation="WADQKIRKQKQKDTLTPAKNSTSTMPSPNSQISLNLNSLY
 GSINTSPQLSSPTSSDASCSGSPMPPTPSIPMMPPTSDCSPTTQPTSSIPSSIA
 STSPMTSLPLSGFPLINIRDPSEASLMLGNADAAALLKTILDDYRLMEASM
 SMSSPKSDSDGSDGLNSIIDVTYVPICLSRAPVLSNVRPHPLHLTLD
 CDYSAHCDLPCCPLFS"
 BASE COUNT 11789 a 6371 c 6437 g 11935 t
 ORIGIN
 Query Match 41.0%; Score 33.2; DB 3; Length 36532;
 Best local similarity 64.1%; Pred. No. 2.6;
 Matches 50; Conservative 0; Mismatches 28; Indels 0; Gaps 0;

SOURCE
 ORGANISM
 Caenorhabditis elegans.
 Caenorhabditis elegans
 Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
 Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REMARK
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 COMMENT
 Submitted (27-OCT-1999) Nematode Sequencing Project, Sanger
 Institute, Hinxton, Cambridge CB10 1SA, England and Department of
 Genetics, Washington University, St. Louis, MO 63110, USA. E-mail:
 jes@sanger.ac.uk or tw@nematode.wustl.edu
 On May 14, 2001 this sequence version replaced gi:5730145.
 Coding sequences below are predicted from computer analysis, using
 predictions from GeneFinder (P. Green, U. Washington), and other
 available information.
 Current sequence finishing criteria for the C. elegans genome
 sequencing consortium are that all bases are either sequenced
 unambiguously on both strands, or on a single strand with both a
 dye primer and dye terminator reaction, from distinct subclones.
 Exceptions are indicated by an explicit note.
 See: a graphical representation of this sequence and its analysis
 for: [http://wormbase.sanger.ac.uk/perl/ace/elegans/seq/sequence?](http://wormbase.sanger.ac.uk/perl/ace/elegans/seq/sequence?name=y51H4A)
 name=y51H4A
 IMPORTANT: This sequence is NOT necessarily the entire insert of
 the specified clone. It may be shorter because we only sequence
 overlapping sections once, or longer because we arrange for a small
 overlap between neighbouring submissions.
 This sequence is the entire insert of clone Y51H4A. The true left
 end of clone Y43D4A is at 193825 in this sequence. The start of
 this sequence (1..115) overlaps with the end of sequence AL713992.
 The end of this sequence (269514..269613) overlaps with the start
 of sequence AL13246.
 Location/Qualifiers
 1..269619
 /organism="Caenorhabditis elegans"
 /db_xref="taxon:6239"
 /chromosome="IV"
 /clone="Y51H4A"
 complement(join(2293..2346,2393..2610,2668..2758,
 3398..3694,5250..5357))
 /gene="Y51H4A.1"
 complement(join(2293..2346,2393..2610,2668..2758,
 3398..3694,5250..5357))
 /gene="Y51H4A.1"
 /note="predicted using GeneFinder"
 /codon_start=1
 /protein_id="CAB61134.1"
 /db_xref="GI:6434492"
 /db_xref="SPTREMBL:Q9NAE5"
 /translation="MLVLYLVILFTLLFPYFKLLIFGKQIVENSIEKINSNGIS
 LAENSKNLEKSSQSDNENAELEIHTAAAPENYLEEIOIAKVEVEDEDEK
 PAGSPKSTSCISSEILSEOSIGTCAKLCFNPRLPLVGRHDAEYCCRRVARPOFO
 NFNCRFESSIPRSPSPFDPATSSASLSPTRTPPIVVRMMSSEERSLHKTEK
 RKSDIDLRLRIQSAKRHRFRSEHLQOVQHHHDY"
 complement(join(12157..12228,13962..14312,16459..16593,
 17111..17131))
 /gene="Y51H4A.3"
 complement(join(12157..12228,13962..14312,16459..16593,
 17111..17131))
 /gene="Y51H4A.3"
 /note="contains similarity to Pfam domain: PF00071 (Ras
 family), Score=317.2, E-value=6.4e-92, N=1
 cDNA EST yk60d10.5 comes from this gene
 cDNA EST yk10347.5 comes from this gene
 cDNA EST yk77f3.5 comes from this gene

```

cdna EST yk23647.5 comes from this gene
cdna EST yk43517.5 comes from this gene
cdna EST yk53112.5 comes from this gene
cdna EST yk5298.5 comes from this gene
cdna EST yk651f5.5 comes from this gene
cdna EST yk671f3.5 comes from this gene
cdna EST yk565h1.5 comes from this gene"
/codon_start=1
/protein_id="CAB63379.1"
/db_xref="GI:6580310"
/translation="MAAIRKILVYIGDCAAGCTGLIYFSKPKQPDVYVYFENVA
DIEDVQKVELAMDPTAGQEDYDLRLPISTPEPTDYLIMCFSDSDSLNPEKTPF
VRHPEPNVPIILVKGKRLDRSDPQVRLRKAQKQPVAPPEQGRALAEQIGAFVLECS
AKTQDIEVEPEKIAQVALQDQKKKKSCML"
complement(join(21593. .21756,22186. .22384,22466. .22675))
/gene="Y514A.2"
complement(join(21593. .21756,22186. .22384,22466. .22675))
/gene="Y514A.2"
complement(join(21593. .21756,22186. .22384,22466. .22675))
/gene="Y514A.2"
/codon_start=1
/protein_id="CAB63380.1"
/db_xref="GI:6580311"
/translation="MKSVGGSGSRLLILSRRHLSLPLDGGREPSLESHDPHY
IHGCLIAAREPQKQKGLADPRFOENSTTALLCGRVPLIARKLSFPHPIGARRS
VDNGKMKRRGAKALITLOFSSFEFLDKMTKNEVUSYFSSJFCYFPKTPFR
LGLDIPALMKLSVRAAERERRAARTIV"
complement(join(31618. .31755,32295. .32402,33420. .33614,
33718. .33834))
/gene="Y514A.4"
complement(join(31618. .31755,32295. .32402,33420. .33614,
33718. .33834))
/feature="predicted using GeneFinder
cdna EST yk282h7.3 comes from this gene
cdna EST yk340d3.3 comes from this gene
cdna EST yk4479g1.3 comes from this gene
cdna EST yk282h7.5 comes from this gene
cdna EST yk340d3.5 comes from this gene
cdna EST yk4479g1.5 comes from this gene
cdna EST yk486f5.5 comes from this gene
cdna EST yk573b7.3 comes from this gene
cdna EST yk573b7.5 comes from this gene"
/codon_start=1
/protein_id="CAB61136.1"
/db_xref="GI:6434494"
/translation="MMPTDILQSLAASRACDSDSIAFNQNTQNPYGAGAFPKNSAA
TALPFAQHPAYAPAHVATISFSPSSSTSTSSQNOQSGSSGSKKRPVPPEKQK
DEVALDRRRRNNEARAKRSREKSKYKDOONSVYVTLERENCLRLVYQLOLQNESMR
OHLILQNGDITDMPIMANPSP"
complement(join(50748. .50898,51588. .51977,52741. .52920,
53245. .53366,53557. .53611,54196. .54274,54700. .54854,
55009. .55055,57238. .57575))
/gene="Y514A.5"
complement(join(50748. .50898,51588. .51977,52741. .52920,
53245. .53366,53557. .53611,54196. .54274,54700. .54854,
55009. .55055,57238. .57575))
/gene="Y514A.5"
complement(join(50748. .50898,51588. .51977,52741. .52920,
53245. .53366,53557. .53611,54196. .54274,54700. .54854,
55009. .55055,57238. .57575))
/feature="predicted using GeneFinder
contains similarity to Pfam domain: PFO164 (Lipase),
score=244.3, E-value=5.4e-70, N=1"
/codon_start=1
/protein_id="CAB61137.1"
/db_xref="GI:6434495"
/translation="MERLNMQCLYLHLFLVYCCWIVGTSAIGRAEYTSHEKTIHL
DKDAIRENSDHLKMYTQWCDSTLSKYSFQILHSAALRIFIDSAVGSALISKL
LEEDYKNTIKKORNEFKORQAPREYVEVEKEFKEKTLLELEFACENFCIRSNIE
FNSAANLEATITVEFLIEMQEDLEASKCDKCIASGHNMCVTYNEGCFSCAIS
KSLNCELPDKAHAYDNDNPRKTYIPLASALINPNPEKCEKRLPLPTLISKFSANCS
YVDFEFLPMNAGLOODITQIKKYKDYDGLVGHGSHGALASVAASVAVHTGIFPSDK

```

```

VKLVTLGQPTGGDYATYMHDKNPPEFRIVHNRDLVAHIIPDGDADKILEHHSEWY
MNNKSTIEPHYICAEADGFYCSSLRELDITNADHLITGVVTPGMOTKC"
join(61826..61986,62058..62127)
/gene="y51H4A.6"
join(61826..61986,62058..62127)
/gene="y51H4A.6"
/codon_start=1
/protein_id="CAB63381.1"
/db_xref="GI:6580312"
/translation="MHRNYFKKKKNLIFSEKQOOSVEKYGGDWVHHFTTANARTEFGK
IFGFIFFRNLSIETRETPTAKESGLDHNNY"
complement(join(63978..64270,64412..64705,66455..66576,
66629..66970,67242..67534,68927..69055,69102..69243,
69290..69561,70362..70488,70898..71079))
/gene="y51H4A.7"
complement(join(63978..64270,64412..64705,66455..66576,
66629..66970,67242..67534,68927..69055,69102..69243,
69290..69561,70362..70488,70898..71079))
/gene="y51H4A.7"
/note="predicted using GeneFinder
contains similarity to Pfam domain: PF01175 (Urocanase),
score=350.6, E-value=5.7e-102, N=1
cdna EST yk156d6.5 comes from this gene
cdna EST yk155e6.5 comes from this gene
cdna EST yk155e6.3 comes from this gene
cdna EST yk259b10.3 comes from this gene
cdna EST yk259b10.5 comes from this gene"
/codon_start=1
/protein_id="CAB61139.1"
/db_xref="GI:6434497"
/db_xref="SPTRMBL:Q9NAE2"
/translation="MOSSDRFVVVKDNLEPEDENHONREFYSGDLLGKEHTTWKG
EEKNIHEKSFIDQNNMNIPSLFVDPSFLLEHPTEORAKNVAAHPKRCNTLOTCK
ASALIIINLNLRKRAVOPOELVTYVGNGCVSNMTOPRLVRYLTMTDHOVLIX
MGALIRNALRILPKEHVLLATEFEAEELNTGYHYIKEMPNPDLAPVSEIGACHEQ
LAATIIIMNLNKRKRAVOPOELVTYVGNGCVSNMTOPRLVRYLTMTDHOVLIX
SGPLGLFPSTPDSPRTVTINGMMITPYSTRKELYDKFYALGYTOYGMTAGSCFYIGP
OGIVHGTTITVLNAGRMDLSDLAGKVFTAGLGMSGAOPKAKTAGCTGIVTAETS D
TALKRHQQGMLDVYSKDLEBEIVMIKEYERKEEALSIGLVAVDMLELAEPEDL
VELSDQTSLNHPLEGFYPAGLTFEQSONMMDSVPKFEKLVSNSLIIRIAIDRIA
AKGYFMWDGNFALEECORAGANLIRRDADDOSFRSYPMOODIMGIFPMGGPFMW
VCSGRPREDLRTDOTCAKTIIDELKPDPVEYKOOLNMKKRIEAEKKKLIVGSOA
RIKLSADRGRYALASAFNELVKSQSAIALIVISRDHHDSGTSPRETSINYDGSNF
TADMAVONCIGDSFRGATWVALHNGGVGNDVINSGFGLVLDGSSDAARAEGMLNW
DVPNGVTRRSWSGNAKAOEAIOREAKEOVDELRYTLPPVEADELLKLKF"
80385..80457
/gene="y51H4A.t1"
80385..80457
/gene="y51H4A.t1"

```

Query Match 41.0%, Score 33.2; DB 3; Length 269619;
Best Local Similarity 64.1%; Pred. No. 3.1;
Matches 50; Conservative 0; Mismatches 28; Indels 0; Gaps 0;

Oy	1	GGACCTCCAGCACACC	CGGCCCCAGAGAAGTG	GCGCTCCTGTGCACCA	GGTTTA	60
Dbl30146	GGACCA	CACGAGCACTTCG	GAGCTCCAGACA	CAAAAAGGACCTT	CAGGACCA	130087
Oy	61	CCACATATACAG	GAGAA	78		
Dbl30086	CCAGACA	CAATCTGAGGA	130069			

RESULT 9						
LOCUS AF282902	5851 bp	mRNA	linear	INV 17-Dec-2000		
DEFINITION Hydra vulgaris type IV collagen alpha 1 chain precursor, mRNA,						
complete cds.						
ACCESSION AF282902						
VERSION AF282902.1	GI:11875611					
KEYWORDS						
SOURCE Hydra vulgaris.						
ORGANISM Hydra vulgaris						

REFERENCE 1 (bases 1 to 6512)
 AUTHORS Muthukumar, G., Blumberg, B. and Kurkinen, M.
 TITLE The complete primary structure for the alpha 1-chain of mouse collagen IV. Differential evolution of collagen IV domains
 JOURNAL J. Biol. Chem. 264 (11), 6310-6317 (1989)
 MEDLINE 89197932
 PUBMED 2703490

COMMENT On Oct 8, 1994 this sequence version replaced gi:340547.
 FEATURES

source location/Qualifiers
 1..6512
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 /tissue_type="parietal endoderm"
 /dev_stage="13.5 day old embryo"
 1..6512
 /gene="Col4a-1"
 132..5141
 /gene="Col4a-1"
 /codon_start=1
 /product="alpha-1 type IV collagen"
 /protein_id="AA50292.1"
 /db_xref="GI:556297"
 /translation="MGPRLSVWLILFLFAALLHERRSRAAKGGCGSSGCGKCDCHY
 KGKGERGLPGLGVIGIPGPMGMPGPPGPGKGDAGEPGLPGTKRGPPGAAAGY
 GNPLRGIRPGDDGPPPGPIRGCMNTKERPLPGPLPGTSNGRPPGLPEMKDPG
 EILHIVPGLTKGERPPIIPGMPSPGLPGLGVGPPGFTGPPGPPGLPPEKDG
 QMGSSFDGPKDKDEQVSGPPVPGVAGVKEGDFAPTEGKGAGEGFPVPGYE
 KGEPPKQPRKPKDKGKERGSGFIQGDVSGYPLPGRGPPQKGEKSGALPLPGTV
 IGTMPLGEKGDGRYGAAPLRGEPGPKGFPPTGGGPPGPPPTGQAAGAPGPPGGE
 KGDDFPFVSLRPGSRDAGAPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG
 PGLTNGGKQKQKESCLACTEGLRPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG
 EGLPGQSPGLIGPAGKGEPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG
 PGLPGKPSPESTIGLKERPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG
 GPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG
 AVGGPIGFPPLPGLPKVGLPGLPGLPGLPGLPGLPGLPGLPGLPGLPGLPGLPGL
 LKGGPGLGIPGTGEEKSGISGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG
 PGIGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG
 SGLPGLRQOSTPVYPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG
 GPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG
 TPVPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG
 PPSGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG
 TPGMPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG
 AGGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG
 VKGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG
 GPKDDKGNQMPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG
 LPLQGVGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG
 PKGGQGVGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG
 PGTSPVHGLVTRHSOTDPLCPGPKILYHGYSILLYOGNRAHGDISTGASCL
 RKSTMPLEFCNNVNCNFAASNDYSLSTPEPPMAMAPISGNDITPTISRCAYCE
 APANVMAVHSOTIQPCPNMGSSLIWISFVHMHSAAEGSGALASPGSCLEEFRS
 APFIECHGRGTCNYANYSFWLATIERSEMFKPTPTSLAKAGELTRHVSRCQVCMRR
 T"

polyA_signal 5858..5863
 /note="putative"
 /gene="Col4a-1"
 /note="putative"
 polyA_site 6512
 /gene="Col4a-1"
 BASE COUNT 1553 a 1798 c 1950 g 1211 t
 ORIGIN
 Query Match 39.5%; Score 32; DB 10; Length 6512;
 Best Local Similarity 62.5%; Pred. No. 5.7;
 Matches 50; Conservative 0; Mismatches 30; Indels 0; Gaps 0;
 Oy 1 GGACCTCCAGCACACCCAGGCCCCCAAGAAGTGGGCTCTCTGTGACAGGTTTA 60
 db 2694 GGACACTCAGGCGCTCCCTGCTTCTTGACACAGGAGGACACTTGAGTCCAGGGTTC 2753
 Oy 61 CCACATATACAGAGAAAT 80
 db 2754 CCAGGTTCTAAGGTAAAT 2773
 RESULT 14
 HSAJ4862 15143 bp DNA linear PRI 10-MAR-2000
 LOCUS Homo sapiens partial MUC5B gene, exon 1-29.
 DEFINITION
 ACCESSION AJ004862
 VERSION AJ004862.1 GI:4038586
 KEYWORDS MUC5B gene; mucin.
 SOURCE Homo sapiens.
 ORGANISM Homo sapiens.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 1 (bases 1 to 15143)
 AUTHORS Desseyn, J. L., Buisine, M. P., Porchet, N., Aubert, J. P. and Laine, A.
 TITLE genomic organization of the human mucin gene MUC5B. cDNA and
 genomic sequences upstream of the large central exon
 JOURNAL J. Biol. Chem. 273 (46), 30157-30164 (1998)
 MEDLINE 99023932
 PUBMED 9804771
 REFERENCE 2 (bases 1 to 15143)
 Laine, A.
 TITLE Direct Submission
 JOURNAL Submitted (23-MAR-1998) Laine A., Laboratoire G Biserete, INSERM
 U377, Place de Verdun, 59045 Lille cedex, FRANCE
 FEATURES
 source location/Qualifiers
 1..15143
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="11"
 /map="p15.5"
 1..15143
 /gene="MUC5B"
 <1..46
 /gene="MUC5B"
 /number=1
 <1..36
 /gene="MUC5B"
 join(37..46,511..582,921..1182,1337..1451,1614..1704,
 1979..2085,2872..3073,3477..3602,3997..4114,4311..4449,
 4795..4905,5243..5312,5732..5869,6308..6469,6760..6854,
 6953..7079,7322..7577,8101..8156,8504..8604,8828..8879,
 9373..9511,9610..9720,10696..10872,11230..11469,
 13173..13329,13742..13870,14102..14246,14347..14390,
 14478..>14675)
 /gene="MUC5B"
 /codon_start=1
 /product="mucin"
 /protein_id="CAA06167.1"
 /db_xref="GI:4038587"
 /translation="MDGAPTSPPRRVSVFPVTPFSLSPLNPAHNGRCSTMGDF
 HYKFDGVPFRPGLCNVFSBHCAAAEDENVOALRRLVGSRRPVYRVYIKAGSLV

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 19, 2003, 23:48:42 : Search time 1001.14 Seconds
(without alignments)
6406.126 Million cell updates/sec

Title: US-09-997-610-1_COPY_53_448

Perfect score: 396
Sequence: 1 gctgactccagcacacc.....cttggtgtcttaagga 396

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estcnu:*
5: em_estcov:*
6: em_estpl:*
7: em_estro:*
8: em_hic:*
9: gb_est1:*
10: gb_est2:*
11: gb_hic:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_tod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Query Length	DB ID	Description
1	113	28.5	499	14	BM967732 LM24HM013
2	76.4	19.3	504	12	BE757275 211668 MA
3	76.4	19.3	526	12	BE757276 211669 MA
4	61.8	15.6	612	17	AG088117 Pan t10g1
5	59.6	15.1	538	9	AA777621 z195a07.s
6	58.6	14.8	440	10	BE062167 RCI-BT025

c 7	58.6	14.8	475	17	B36584	HS-1041-A1-
c 8	58.6	14.8	537	10	BE079777	RC6-BT062
c 9	58.6	14.8	563	10	BE079876	RC6-BT062
c 10	58.6	14.8	678	10	AM813783	RC3-ST019
c 11	58.6	14.8	727	10	AV731140	AV731140
c 12	58.6	14.8	2615	10	BE420422	BE420422
c 13	58	14.6	345	17	AQ012504	CIT-HSP-2
c 14	58	14.6	342	12	BF881529	AV1-ET018
c 15	58	14.6	670	17	AC093990	Pan t10g1
c 16	57.8	14.6	397	13	BT180953	BT180953
c 17	57	14.4	425	10	AM517269	AM517269
c 18	57	14.4	466	9	AA372508	AA372508
c 19	57	14.4	620	17	AC019007	AC019007
c 20	57	14.4	666	17	AG141320	AG141320
c 21	57	14.4	683	17	AC019080	AC019080
c 22	57	14.4	1036	13	BM462674	BM462674
c 23	56.8	14.3	577	10	AV716883	AV716883
c 24	56.6	14.3	662	10	BE390081	BE390081
c 25	56.4	14.2	284	9	AA360873	AA360873
c 26	56.4	14.2	461	12	BF828364	BF828364
c 27	56.4	14.2	461	12	BF829000	BF829000
c 28	56.4	14.2	545	10	BE279213	BE279213
c 29	56.4	14.2	609	10	BE389158	BE389158
c 30	56.4	14.2	667	17	AC080755	AC080755
c 31	56.4	14.2	668	17	AC080765	AC080765
c 32	56.4	14.2	698	12	BE729966	BE729966
c 33	56.4	14.2	705	17	AC052122	AC052122
c 34	56.4	14.2	721	12	BE728927	BE728927
c 35	56.4	14.2	795	12	BF025976	BF025976
c 36	56.4	14.2	800	12	BF203746	BF203746
c 37	56.4	14.2	885	12	BF304683	BF304683
c 38	56.4	14.2	885	12	BF304683	BF304683
c 39	56.4	14.2	1080	14	BQ422247	BQ422247
c 40	56	14.1	315	9	AA719635	AA719635
c 41	56	14.1	645	17	AC014562	AC014562
c 42	55.8	14.1	360	10	AM898344	AM898344
c 43	55.4	14.0	364	17	AO665181	AO665181
c 44	55.4	14.0	388	17	B78843	CYT-HSP-734
c 45	55.4	14.0	453	17	AO393565	AO393565

ALIGNMENTS

RESULT 1
LOCUS BM967732 499 bp mRNA linear EST 20-MAR-2002
DEFINITION LM24HM0134 Bos taurus LM-24-HW cDNA library Bos taurus cDNA clone
ACCESSION BM967732
VERSION BM967732.1 GI:19561919
KEYWORDS EST.
SOURCE cow.
ORGANISM Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovidae; Bovinae; Bos.
REFERENCE 1 (bases 1 to 499)
Yoon,D.H., Jang,Y.S., Kim,T.H., Park,E.W., Lee,H.K., Chung,E.R.,
Sun,S.S. and Cheong,I.C.
Gene Expression Profiling of the Bovine skeletal muscle
Unpublished (2002)
CONTACT: Dr. Du-Hak Yoon
National Livestock Research Institute, RDA
564 Omoekchun-dong, Suwon, 441-350, Korea
Tel: 82 31 290 1593
Fax: 82 31 290 1792
Email: dhyoon@rda.go.kr
Insert Length: 499 Std Error: 0.00
Seq primer: CAGGAACACGTATGAC
POLYA-NO.
FEATURES
source location/Qualifiers
L. 499

```
/organism="Bos taurus"
/db_xref="taxon:9913"
/clone="LM-24-HW-011-34 (5')"
/sex="six males mixed"
/tissue.type="longissimus dorsi"
/cell.type="myocyte"
/dev_stage="24 months old"
/lab_host="X11-BlueMF strain"
/note="Organ: skeletal muscle; Vector: Uni-ZAPPR; Site_1:
ECORI; Site_2: Xho I"

BASE COUNT      129 a      139 c      131 g      100 t
ORIGIN

Query Match      28.5%; Score 113; DB 14; Length 499;
Best Local Similarity 64.7%; Pred. No. 5.8e-21;
Matches 200; Conservative 0; Mismatches 105; Indels 4; Gaps 2;

QY      2 CTGACCTCGACGACCGCCGACGAGGAGGAGTGGGGCTCTGTCGACCCAGGTT 61
      11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11
Db      181 CGGGGCCACGACGACGACGAGGCGCTCCAGGTATAGAGGCGCTCCAGGTATAGAGGAA 240
QY      62 TACCACAAATATACAGGA---GAATATAGTGAATGACAAATGCCCTGCTGATATAG 118
      11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11
Db      241 TACCAGGTTTGCAGAGTCGCGCCGGGAACTCCAGGACCAAGTGAATGCCCATGCCACA 300
QY      119 AAAGGTACGCTTACTGTGTAAGCTCAGTGAAACTCTCTCTTCTTCAAGCCATCA 178
      11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11
Db      301 GACAGTCCGCTTCACTCACTAAGCTCAGTGCGCAGTTCCTCCCTTCAAGACCTGTGC 360
QY      179 TC-TTCACAGGGGCTGTCATATGCCAGAGGATTTAAGGAGGCGCATGGAGTCTT 237
      11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11
Db      361 CTTTTCACAGAGGCTCGTCAATGCCCCAGAAAGACTTACAGAGGACACCTGGGCTTTC 420
QY      238 GCTTGCAGGGTGGCTGGGAATTACTCAAGCTTTGATGTTGATGTCATCATTTGCA 297
      11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11
Db      421 ACATGCAGGGTGCAGGAATTAACATTTCTTTCTATGTGATCTCATCTGCAAG 480
QY      298 GTGATATTT 306
      11 11 11 11 11
Db      481 GTGACTGTT 489

RESULT 2
LOCUS      BE757275      504 bp      mRNA      linear      EST 25-APR-2001
DEFINITION      211668 MARC 2B0V Bos taurus cDNA 5', mRNA sequence.
ACCESSION      BE757275
VERSION      BE757275.1 GI:10171267
KEYWORDS      EST.
SOURCE      cow.
ORGANISM      Bos taurus
      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
      Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
      Bovidae; Bovinae; Bos.
REFERENCE      1 (bases 1 to 504)
AUTHORS      Smith,T.P., Grosse,W.M., Freking,B.A., Roberts,A.J., Stone,R.T.,
      Casas,E., Wray,J.E., White,J., Cho,J., Fahrenkrug,S.C., Bennett,
      G.L., Heaton,M.P., Laegreid,W.W., Rohrer,G.A., Chitko-McKown,C.G.,
      Pereira,G., Holt,I., Karanycheva,S., Liang,F., Quackenbush,J. and
      Keeler,J.W.
      Sequence evaluation of four pooled-tissue normalized bovine cDNA
      libraries and construction of a gene index for cattle
      Genome Res. 11 (4), 626-630 (2001)
JOURNAL      MEDLINE
COMMENT      21180013
      Contact: Smith TPL
      USDA, ARS, US Meat Animal Research Center
      PO Box 166, Clay Center, NE 68933-0166, USA
      Tel: 402 762 4366
      Fax: 402 762 4390
      Email: smith@email.marc.usda.gov
      Single pass sequencing. Bases called and alt_trimmed with phred
      v0.980904.e. Vector identified by cross_match with the -mnscore 18
```

```
and -mismatch 12 options.
PCR Primers
FORWARD: AGGAACACCTATGACCAT
BACKWARD: GTTTCACGTCACGACG
Plate: 63 row: G column: 13
Seq primer: ATTAGGTGACTATAG.
location/Qualifiers
1..504
/organism="Bos taurus"
/db_xref="taxon:9913"
/clone_lib="MARC 2B0V"
/tissue.type="pooled"
/lab_host="DH10B"
/note="Vector: pCMV SPORT6; Site_1: NotI; Site_2: SalI;
Library made from pooled tissue from testis, thymus,
semitendinosus muscle, longissimus muscle, pancreas,
adrenal, and endometrium."

BASE COUNT      99 a      131 c      146 g      128 t
ORIGIN

Query Match      19.3%; Score 76.4; DB 12; Length 504;
Best Local Similarity 61.6%; Pred. No. 1.1e-10;
Matches 122; Conservative 0; Mismatches 76; Indels 0; Gaps 0;

QY      123 GTGACCTTACTGTGAAGCTCACTGGAACACTTCCTCTTCAAGCCATCATCTT 182
      11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11
Db      303 GTGCGCTTTCCCTGATGAGTGAAGTGGGCTTCCAGGACCTCCAGCATTTGCTT 362
QY      183 CACAGGGGCTCTGTACAAATGCCAGAGGATTTAAGAGGCGCATGGAGCTTTGCTTG 242
      11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11
Db      363 CAGAGAGTCTGTACAAACATCAGGGCCACTTGACCCGCACTGTGTGTCAGCTG 422
QY      243 CAGGGTCCCTGGGAATTACTCAAGCTTTGATGTTGACCTGCATTCATTCAGAGTGAA 302
      11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11
Db      423 CAGCGTCCCTGGTGTGACACACTTTGCTTGAATGAGTTGTTTCAGAGTCTGTCAA 482
QY      303 TATTTGGCTAATGAGGAA 320
      11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11
Db      483 GTGGGTCTAATGCGGAA 500

RESULT 3
LOCUS      BE757276      526 bp      mRNA      linear      EST 25-APR-2001
DEFINITION      211669 MARC 2B0V Bos taurus cDNA 5', mRNA sequence.
ACCESSION      BE757276
VERSION      BE757276.1 GI:10171268
KEYWORDS      EST.
SOURCE      cow.
ORGANISM      Bos taurus
      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
      Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
      Bovidae; Bovinae; Bos.
REFERENCE      1 (bases 1 to 526)
AUTHORS      Smith,T.P., Grosse,W.M., Freking,B.A., Roberts,A.J., Stone,R.T.,
      Casas,E., Wray,J.E., White,J., Cho,J., Fahrenkrug,S.C., Bennett,
      G.L., Heaton,M.P., Laegreid,W.W., Rohrer,G.A., Chitko-McKown,C.G.,
      Pereira,G., Holt,I., Karanycheva,S., Liang,F., Quackenbush,J. and
      Keeler,J.W.
      Sequence evaluation of four pooled-tissue normalized bovine cDNA
      libraries and construction of a gene index for cattle
      Genome Res. 11 (4), 626-630 (2001)
JOURNAL      MEDLINE
COMMENT      21180013
      Contact: Smith TPL
      USDA, ARS, US Meat Animal Research Center
      PO Box 166, Clay Center, NE 68933-0166, USA
      Tel: 402 762 4366
      Fax: 402 762 4390
      Email: smith@email.marc.usda.gov
      Single pass sequencing. Bases called and alt_trimmed with phred
      v0.980904.e. Vector identified by cross_match with the -mnscore 18
      and -mismatch 12 options.
      PCR Primers
```


and Eco RI sites of the modified pT73 vector. Library went through one round of normalization. Library constructed by Bento Soares and M.Fatima Bonaldo."

BASE COUNT

134 a 148 c 81 g 175 t

ORIGIN

Query Match 15.1%; Score 59.6; DB 9; Length 538;
Best Local Similarity 70.2%; Pred. No. 5.7e-06;
Matches 80; Conservative 0; Mismatches 34; Indels 0; Gaps 0;

QY 283 CTGCATCATTCGACAGCTGAATTTGGCTAATGAGCAACAAATTTGGCTAATTAAGCA 342
DB 178 CTGCCCTAGAGATGTGTGACACTTTCATTGAGAGAGCTGATTCGGGATATCGACAGAA 119
QY 343 GAAATTTCTAAGCAGCAAGCATTCAGAGGTCGCTGGCTGTTAAAGCA 396
DB 118 GAAATTTCTAAGCAGCAAGCATTCAGAGGTCGCTGGCTGTTAAAGCA 65

RESULT 6

BE062167/c

LOCUS BE062167 440 bp mRNA linear EST 09-JUN-2000
DEFINITION RC1-BT0254-220300-019-c05 BT0254 Homo sapiens cDNA, mRNA sequence.
ACCESSION BE062167
VERSION BE062167.1 GI:8406817
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
1 (bases 1 to 440)
AUTHORS Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R.,
Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F.,
Goldman, G.H., Carvalho, A.F., Matsukuma, A., Bala, G.S., Simpson, D.H.,
Brunstein, A., deoliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare,
M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and
Simpson, A.J.

TITLE Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags

JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE 20202663
COMMENT Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001

Email: asimpson@ludwig.org.br
This sequence was derived from the RAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=tl2-RC1-BT0254-220
300-019-c05&tl=2000-03-22&tl=1)
Seq primer: puc 18 forward
High quality sequence stop: 440.
Location/Qualifiers

FEATURES

source

1. 440
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="BT0254"
/dev_stage="Adult"
/note="Organ: Breast; Vector: puc18; Site.1: SmaI; Site.2:
SmaI; A mini-library was made by cloning products derived
from ORESTES PCR (U.S. Letters Patent application No. 196
716 - Ludwig Institute for Cancer Research) profiles
into the pUC 18 vector. Reverse transcription of tissue
mRNA and cDNA amplification were performed under low
stringency conditions."

BASE COUNT 143 a 119 c 66 g 112 t

ORIGIN

Query Match 14.8%; Score 58.6; DB 10; Length 440;
Best Local Similarity 75.3%; Pred. No. 1e-05;

Matches 73; Conservative 0; Mismatches 24; Indels 0; Gaps 0;

QY 300 GAATTAATTTGGCTAATGAGCAAGCAAAATTTGGCTAATTAAGCAAAATTTCTAAGCAGCA 359
DB 237 GAACCTTTGAACCTTGAGAGATGATTTATGCTGCGCAGAGAAATTTCTAAGCAGCA 178
QY 360 AACCAATTCAGAGGTCGCTGGCTGCTGTTAAAGCA 396
DB 177 AACCAATTCAGAGGTCGCTGGCTGCTGTTAAAGCA 141

RESULT 7

B36584/c

LOCUS B36584 475 bp DNA linear GSS 17-OCT-1997
DEFINITION HS-1041-A1-D06-MF.abi CTF Human Genomic Sperm Library C Homo
sapiens genomic clone Plate=CT 823 Col=11 Row=G, DNA sequence.
ACCESSION B36584
VERSION B36584.1 GI:2535953
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
1 (bases 1 to 475)
AUTHORS Traicoff, R., Abajian, C., Blanchard, A., West, A. and Hood, L.E.
Construction of a Characterized Clone Resource for Genomic
Sequencing: Generation and Preliminary Analysis of 20,000 Sequence
Tagged Connectors
Unpublished (1997)
COMMENT Contact: Mahairas G., Zackrone KD, Hood L
University of Washington
Seattle, WA 98195, USA
Tel: (206) 616-8744
Fax: (206) 685-7301
Email: kzackrone@u.washington.edu

TITLE

JOURNAL Tagged Connectors
Unpublished (1997)
COMMENT Contact: Mahairas G., Zackrone KD, Hood L
University of Washington
Seattle, WA 98195, USA
Tel: (206) 616-8744
Fax: (206) 685-7301
Email: kzackrone@u.washington.edu

Sequence Tagged Connector
Plate: CT 823 row: G column: 11
Class: BAC ends
High quality sequence stop: 475.
Location/Qualifiers

FEATURES

source

1. 475
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="Plate=CT 823 Col=11 Row=G"
/note="Organ: sperm; Vector: pBeloBAC11; BAC Clones in
E-Coli DH10B"

BASE COUNT 116 a 126 c 86 g 146 t 1 others

ORIGIN

Query Match 14.8%; Score 58.6; DB 17; Length 475;
Best Local Similarity 75.3%; Pred. No. 1e-05;
Matches 73; Conservative 0; Mismatches 24; Indels 0; Gaps 0;

QY 300 GAATTAATTTGGCTAATGAGCAAGCAAAATTTGGCTAATTAAGCAAAATTTCTAAGCAGCA 359
DB 351 GAACCTTTGAACCTTGAGAGATGATTTAGGCTATCGACGAGAGAAATTTCTAAGCAGCA 292
QY 360 AACCAATTCAGAGGTCGCTGGCTGCTGTTAAAGCA 396
DB 291 AACCAATTCAGAGGTCGCTGGCTGCTGTTAAAGCA 255

RESULT 8

BE079777/c

LOCUS BE079777 537 bp mRNA linear EST 12-JUN-2000
DEFINITION RC6-BT0627-140200-011-A05 BT0627 Homo sapiens cDNA, mRNA sequence.
ACCESSION BE079777
VERSION BE079777.1 GI:8470060
KEYWORDS EST.
SOURCE human.

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
REFERENCE 1 (bases 1 to 537)
AUTHORS Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R.,
Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F.,
Goldman, G.H., Carvalho, A.F., Matsukuma, A., Bata, G.S., Simpson, D.H.,
Brunstein, A., deOliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare,
M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and
Simpson, A.J.
TITLE Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE 20202663
COMMENT Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the PAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=4t2-RC6-BR0627-140
200-011-A054t3-2000-02-14&tl=1)
Seq primer: puc 18 forward
High quality sequence start: 69
High quality sequence stop: 537.
Location/Qualifiers
1. 537
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="BR0627"
/dev_stage="Adult"
/note="Organ: breast; Vector: puc18; Site_1: SmaI; Site_2:
SmaI; A mini-library was made by cloning products derived
from ORESTES PCR (U.S. Letters Patent application No. 196
, 716 - Ludwig Institute for Cancer Research) profiles
into the pUC 18 vector. Reverse transcription of tissue
mRNA and cDNA amplification were performed under low
stringency conditions."

BASE COUNT 163 a 135 c 100 g 137 t 2 others
ORIGIN

Query Match 14.8%; Score 58.6; DB 10; Length 537;
Best Local Similarity 75.3%; Pred. No. 1.1e-05;
Matches 73; Conservative 0; Mismatches 24; Indels 0; Gaps 0;

QY 300 GAATATTGGCTAATGAGAGCAAAATTTGGCTAATAAGAGAAATTTCTAACAGCA 359
DB 315 GAACTTGAACCTCAGAGAGATGATTAAAGTATCTGGTGAGAGAAATTTCTAACAGCA 256
QY 360 AAGCATTCAGAGGTGACTGGTGCTGTAAAGCA 396
DB 255 AAGCATTCAGAGGTGACTGGTGCTGTAAAGCA 219

RESULT 9
BE079876/c 563 bp mRNA linear EST 12-JUN-2000
LOCUS BE079876
DEFINITION RC6-BR0627-220300-012-H08 BR0627 Homo sapiens cDNA, mRNA sequence.
ACCESSION BE079876
VERSION BE079876.1 GI:8470160
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
REFERENCE 1 (bases 1 to 563)
AUTHORS Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R.,
Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F.,
Goldman, G.H., Carvalho, A.F., Matsukuma, A., Bata, G.S., Simpson, D.H.,
Brunstein, A., deOliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare,
M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and
Simpson, A.J.

Brustein, A., deOliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare,
M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and
Simpson, A.J.
TITLE Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE 20202663
COMMENT Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the PAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=4t2-RC6-BR0627-220
300-012-H084t3-2000-03-22&tl=1)
Seq primer: puc 18 forward
High quality sequence start: 32
High quality sequence stop: 563.
Location/Qualifiers
1. 563
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="BR0627"
/dev_stage="Adult"
/note="Organ: breast; Vector: puc18; Site_1: SmaI; Site_2:
SmaI; A mini-library was made by cloning products derived
from ORESTES PCR (U.S. Letters Patent application No. 196
, 716 - Ludwig Institute for Cancer Research) profiles
into the pUC 18 vector. Reverse transcription of tissue
mRNA and cDNA amplification were performed under low
stringency conditions."

BASE COUNT 174 a 144 c 100 g 144 t 1 others
ORIGIN

Query Match 14.8%; Score 58.6; DB 10; Length 563;
Best Local Similarity 75.3%; Pred. No. 1.1e-05;
Matches 73; Conservative 0; Mismatches 24; Indels 0; Gaps 0;

QY 300 GAATATTGGCTAATGAGAGCAAAATTTGGCTAATAAGAGAAATTTCTAACAGCA 359
DB 341 GAACTTGAACCTCAGAGAGATGATTAAAGTATCTGGTGAGAGAAATTTCTAACAGCA 282
QY 360 AAGCATTCAGAGGTGACTGGTGCTGTAAAGCA 396
DB 281 AAGCATTCAGAGGTGACTGGTGCTGTAAAGCA 245

RESULT 10
AW813783/c 678 bp mRNA linear EST 17-MAY-2000
LOCUS AW813783
DEFINITION RC3-ST0197-120200-015-403 ST0197 Homo sapiens cDNA, mRNA sequence.
ACCESSION AW813783
VERSION AW813783.1 GI:7906777
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
REFERENCE 1 (bases 1 to 678)
AUTHORS Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R.,
Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F.,
Goldman, G.H., Carvalho, A.F., Matsukuma, A., Bata, G.S., Simpson, D.H.,
Brunstein, A., deOliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare,
M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and
Simpson, A.J.
TITLE Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE 20202663

COMMENT

Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(<http://www.ludwig.org.br/scripts/gethtml2.pl?lib=at2-RC3-ST0197-120>)
200-015-003&t3=2000-02-12&tt4=1)
Seq primer: puc 18 forward
High quality sequence start: 23
High quality sequence stop: 677.
Location/Qualifiers
1..678

FEATURES

source

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_id="ST0197"
/dev_stage="Adult"
/note="Organ: stomach; Vector: puc18; Site_1: SmaI;
Site_2: SmaI; A mini-library was made by cloning products
derived from ORESPRES PCR (U.S. Letters Patent application
No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the pUC 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."

BASE COUNT

162 a 174 c 108 g 234 t

ORIGIN

Query Match 14.8%; Score 58.6; DB 10; Length 678;
Best Local Similarity 75.3%; Pred. No. 1.2e+05;
Matches 73; Conservative 0; Mismatches 24; Indels 0; Gaps 0;

Db

300 GAATATTGGCTAATAGACGACGAATTTTGGCTATAAGACGAATTTCTTAACACCA 359
||| || | || || | | | | | || |||||||||
250 GAACGTAACTTGAGAGAGATATTTAGCATATGTTGGTGAAGAAGAAATTTCTTAACACCA 191
||||| |||||||

QY

360 AAGCATTCAGAGCGACTGGGCTGCTGTTAAAGCA 396
||||| |||||||

Db

190 AAGCATTCAGAGCGACTGGGCTGCTGTTAAAGCA 154
||||| |||||||

RESULT 11

AV731140 . 727 bp mRNA linear EST 17-OCT-2000
LOCUS AV731140 HTF Homo sapiens cDNA clone HIFAB01.5', mRNA sequence.
DEFINITION AV731140
ACCESSION AV731140
VERSION AV731140.1 GI:10840561
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniota; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
AUTHORS 1 (bases 1 to 727)
Gu,Y., Peng,Y., Song,H., Huang,Q., Yang,Y., Gao,G., Xiao,H., Xu,X.,
Li,N., Qian,B., Liu,F., Qu,J., Gao,X., Cheng,Z., Xu Z., Zeng,L., Xu
,S., Gu,W., Tu,Y., Jia,J., Fu,G., Ren,S., Zhong,M., Lu,G., Hu,R.,
Chen,J., Chen,Z. and Han,Z.
Homo sapiens cDNA HTF clones
Unpublished (2000)
Contact: Zequang Han
Chinese National Human Genome Center at Shanghai
351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai
201203, P. R. China
Tel.: 86-21-50801919(ex.45)
Fax: 86-21-50801922
Email: hanzg@chgc.sh.cn
This clone is available at CHGC in Shanghai.

FEATURES

source

1..727
/organism="Homo sapiens"

[illegible]

DB 553 AAGCATTCAGAGTAACCTGGTGCTGTTAAAGCA 589

RESULT 13
LOCUS A0012504/c 342 bp DNA linear GSS 06-JUN-1998
DEFINITION CIT-HSP-2298M4.TF CIT-HSP Homo sapiens genomic clone 2298M4, DNA sequence.
ACCESSION A0012504
VERSION A0012504.1 GI:3185069
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Mammalia; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE 1 (bases 1 to 342)
Adams,M.D., Rounsley,S.D., Zhao,S., Field,C.E., Bass,S., Linher,K., Golden,K., Berry,K., Granger,D., Suh,E., Wible,C., Shizuya,H., Simon,M. and Venter,J.C.
Use of a random BAC End Sequence Database for Sequence-Ready Map Building (1998)
TITLE Unpublished (1998)
JOURNAL
COMMENT Contact: Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: mdadams@tigr.org
Clones are available from Research Genetics (info@resgen.com). BAC end search page:
http://www.tigr.org/tdb/hungen/bac_end_search/bac_end_search.html.
Seq primer: M13-21
Class: BAC ends.

FEATURES
Location/Qualifiers
source 1..342
/organism="Homo sapiens"
/db_xref="GDB:7153856"
/db_xref="taxon:9606"
/clone="2298M4"
/clone_lib="CIT-HSP"
/sex="Male"
/cell_type="Sperm"
/note="Vector: pBelobAC11; Site_1: HindIII; Site_2: HindIII"

BASE COUNT 88 a 85 c 58 g 111 t

ORIGIN
HindIII

Query Match 14.6%; Score 58; DB 17; Length 342;
Best Local Similarity 100.0%; Pred. No. 1.4e-05;
Matches 58; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 339 GGAAGAAATTTCTAAGCAGCAAGATTCAGAGTGACTGGTGCTGTTAAAGCA 396
|||||
DB 264 GGAAGAAATTTCTAAGCAGCAAGATTCAGAGTGACTGGTGCTGTTAAAGCA 207
|||||

RESULT 14
BE881529/c 345 bp mRNA linear EST 17-JAN-2001
LOCUS OVI-ET0181-031200-546-f03 ET0181 Homo sapiens cDNA, mRNA sequence.
DEFINITION BE881529
ACCESSION BE881529
VERSION BE881529.1 GI:12271655
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Mammalia; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE 1 (bases 1 to 345)
Dies Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Britones,M.R., Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F., Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bala,G.S., Simpson,D.H., Brunstein,A., deoliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare

M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and Simpson,A.J.
Shotgun sequencing of the human transcriptome with ORF expressed sequence tags
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
JOURNAL MEDLINE
COMMENT Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?cl=OVI&ct2=OVI-ET0181-031200-546-f03&ct3=2000-12-03&cl=1)
Seq primer: puc 18 forward
High quality sequence stop: 89.

FEATURES
Location/Qualifiers
source 1..345
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="ET0181"
/dev_stage="Adult"
/note="Organ: lung,tumor; Vector: puc18; Site_1: SmaI; Site_2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the puc 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

BASE COUNT 86 a 96 c 57 g 106 t

ORIGIN
low stringency conditions."

Query Match 14.6%; Score 58; DB 12; Length 345;
Best Local Similarity 100.0%; Pred. No. 1.4e-05;
Matches 58; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 339 GGAAGAAATTTCTAAGCAGCAAGATTCAGAGTGACTGGTGCTGTTAAAGCA 396
|||||
DB 194 GGAAGAAATTTCTAAGCAGCAAGATTCAGAGTGACTGGTGCTGTTAAAGCA 137
|||||

RESULT 15
AG093990/c 670 bp DNA linear GSS 03-NOV-2001
LOCUS Pan troglodytes DNA, clone: PTB-094J23.R, genomic survey sequence.
DEFINITION AG093990
ACCESSION AG093990
VERSION AG093990.1 GI:16645792
KEYWORDS GSS.
SOURCE Pan troglodytes
ORGANISM Pan troglodytes
Mammalia; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Eukaryota; Eutheria; Primates; Catarrhini; Homidae; Pan.
REFERENCE 1
Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
BAC end sequences of library PTB
Unpublished
2 (bases 1 to 670)
Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
Direct Submission
Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center (GSC), Japan
1-7-22 Suenho-chou,Tsukumi-ku, Yokohama, Kanagawa 230-0045, Japan
E-mail:chimpes@sc.riken.go.jp, URL:http://hgp.9sc.riken.go.jp/
Tel:81-45-503-9111, Fax:81-45-503-9170
Clones are derived from the chimpanzee BAC library PTB This BAC end was generated during the Rad process and may have higher chance of

COMMENT

clone tracking errors.

PRIMERS

Sequencing: M13Rev

LIBRARY

Vector : PKS145

R.Site 1 : SacI

R.Site 2 : SacI

Location/Qualifiers

FEATURES

source

1..670

/organism="Pan troglodytes"

/db_xref="taxon:9598"

/clone="PTB-094J23.R"

/sex="male"

/cell_type="lymphoblast"

/clone_lib="PTB Chimpanzee Male BAC library"

155 a 186 c 124 g 205 t

BASE COUNT
ORIGIN

Query Match 14.6% Score 58; DB 17; Length 670;

Best Local Similarity 100.0%; Pred. NO. 1.7e-05;

Matches 58; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 339 GGAAGAATTCTTAAGCAGCAAGCATTCAGAGTGACTGGTGCTGTAAAGGCA 396
|||||
DB 404 GGAAGAATTCTTAAGCAGCAAGCATTCAGAGTGACTGGTGCTGTAAAGGCA 347

Search completed: February 20, 2003, 06:14:33
Job time : 1009.14 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 19, 2003, 18:01:21 ; Search time 31.8656 Seconds
(without alignments)
6329.459 Million cell updates/sec

Title: US-09-997-610-1_COPY_53_448

Perfect score: 396
Sequence: 1 gctgacctccagcacacc.....cttggtgctgtaagga 396

Scoring table:
IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 424239 seqs, 25461826 residues

Total number of hits satisfying chosen parameters: 848478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA:*
1: /cgn2_6/ptodata/1/pubna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubna/PC1_NEW_PUB.seq:*
3: /cgn2_6/ptodata/1/pubna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/1/pubna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/1/pubna/US07_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/1/pubna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubna/US09_NEW_PUB.seq:*
10: /cgn2_6/ptodata/1/pubna/US09_PUBCOMB.seq:*
11: /cgn2_6/ptodata/1/pubna/US10_NEW_PUB.seq:*
12: /cgn2_6/ptodata/1/pubna/US10_PUBCOMB.seq:*
13: /cgn2_6/ptodata/1/pubna/US60_NEW_PUB.seq:*
14: /cgn2_6/ptodata/1/pubna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	396	100.0	1381	9	US-09-997-610-1
2	396	100.0	1731	9	US-09-997-610-5
3	288	72.7	1377	9	US-09-997-610-3
4	288	72.7	1731	9	US-09-997-610-3
5	114.8	29.0	425	10	US-09-960-352-15057
6	93	23.5	415	10	US-09-960-352-9137
7	91.6	23.1	389	10	US-09-960-352-4220
8	82.2	20.8	392	10	US-09-960-352-1786
9	78.4	19.8	410	10	US-09-960-352-1715
10	76	19.2	374	10	US-09-960-352-11516
11	73.2	18.5	447	10	US-09-960-352-3684
12	58.8	14.8	273	10	US-09-960-352-4241
13	58.4	14.7	273	10	US-09-864-761-25080
14	58.4	14.7	519	10	US-09-864-761-8349
15	57	14.4	1946	10	US-09-864-761-2925
16	57	14.4	25603	9	US-09-819-607-3
17	56	14.1	505	10	US-09-864-761-7075
18	55.4	14.0	520	10	US-09-864-761-8643
19	55.4	14.0	542	10	US-09-864-761-8938

20	55.4	14.0	170834	10	US-09-835-232-7	Sequence 7, Appl1
21	55.4	14.0	170834	10	US-09-835-232-7	Sequence 7, Appl1
22	54.8	13.8	525	10	US-09-864-761-8932	Sequence 8932, Ap
23	53.8	13.6	444	10	US-09-864-761-24780	Sequence 24780, A
24	53.8	13.6	484	10	US-09-864-761-837	Sequence 83, Appl
25	53.8	13.6	523	10	US-09-864-761-8787	Sequence 8787, Ap
26	53.8	13.6	543	10	US-09-864-761-8034	Sequence 8034, Ap
27	53.8	13.6	544	10	US-09-864-761-9454	Sequence 9454, Ap
28	53.2	13.4	531	10	US-09-864-761-14775	Sequence 14775, A
29	53.2	13.4	576	10	US-09-864-761-9294	Sequence 9294, Ap
30	53.2	13.4	580	10	US-09-864-761-9802	Sequence 9802, Ap
31	52.6	13.3	552	10	US-09-864-761-8495	Sequence 8495, Ap
32	52.2	13.2	220	10	US-09-864-761-33019	Sequence 33019, A
33	52.2	13.2	516	10	US-09-864-761-7399	Sequence 7399, Ap
34	52.2	13.2	520	10	US-09-864-761-8449	Sequence 8449, Ap
35	52.2	13.2	571	10	US-09-864-761-16494	Sequence 16494, A
36	52.2	13.2	600	10	US-09-864-761-7441	Sequence 7441, Ap
37	52.2	13.2	202001	10	US-09-734-674-3	Sequence 3, Appl1
38	51.6	13.0	473	10	US-09-864-761-2556	Sequence 2556, Ap
39	51.6	13.0	3816	10	US-09-880-107-2298	Sequence 2298, Ap
40	51.6	13.0	42899	10	US-09-740-029-3	Sequence 3, Appl1
41	51.6	13.0	153074	9	US-10-026-188-6	Sequence 6, Appl1
42	51.6	13.0	180557	12	US-10-003-806-6	Sequence 6, Appl1
43	51.6	13.0	180557	12	US-10-003-806-9	Sequence 9, Appl1
44	51.2	12.9	148567	9	US-10-254-869-3	Sequence 3, Appl1
45	51.2	12.9	148567	10	US-09-801-876B-3	Sequence 3, Appl1

ALIGNMENTS

RESULT 1
US-09-997-610-1
; Sequence 1, Application US/09997610
; Patent No. US20020156244A1
; GENERAL INFORMATION:
; APPLICANT: Fox, Brian
; APPLICANT: Holloway, James L.
; TITLE OF INVENTION: ADIPOCYTE COMPLEMENT RELATED PROTEIN
; FILE REFERENCE: ZACRP13
; CURRENT APPLICATION NUMBER: US/09/997,610
; CURRENT FILING DATE: 2001-11-29
; PRIOR APPLICATION NUMBER: US 60/253,924
; PRIOR FILING DATE: 2000-11-29
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 1381
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (2)...(1381)
US-09-997-610-1
Query Match 100.0%; Score 396; DB 9; Length 1381;
Best Local Similarity 100.0%; Pred. No. 3.3e-114;
Matches 396; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
1 GCTGACCTCCAGCACACCCAGGCCCGAGAGAGTGGGGCTCTGTCGACCAAGT 60
|||||
53 GCTGACCTCCAGCACACCCAGGCCCGAGAGAGTGGGGCTCTGTCGACCAAGT 112
|||||
61 TTACCAATATATACAGAGAAATTAAGTGAATGCAAAATGCCCTGCTATATAGAA 120
|||||
113 TTACCAATATATACAGAGAAATTAAGTGAATGCAAAATGCCCTGCTATATAGAA 172
|||||
121 AGTTCAGCTTTACTGTGAAGCTCACTGCAAACTCTCTCTTTCAAGCCATCATC 180
|||||
173 AGTTCAGCTTTACTGTGAAGCTCACTGCAAACTCTCTCTTTCAAGCCATCATC 232
|||||
181 TTCACAGGGGTCTGTATCATGCTCCAGAGAGATTTAAGAGAGCATGGAGTCTTGGCT 240

```
|||||
Db 233 TTCACAGGGGTCTGTCACATGCCCCAGAGGATTTAAGAGGCCATGGAGTCTTTGCT 292
QY 241 TGCAGGGTCTGGGAATTAATCTAGCTCAGCTTGATGTTGAGTGCATCTTGCAGGTG 300
Db 293 TGCAGGGTCTGGGAATTAATCTAGCTCAGCTTGATGTTGAGTGCATCTTGCAGGTG 352
QY 301 AATATTGGCTAATGAGAGCAAAATTTGGCTAATTAAGAGCAAAATTTCTAAGCAGCA 360
Db 353 AATATTGGCTAATGAGAGCAAAATTTGGCTAATTAAGAGCAAAATTTCTAAGCAGCA 412
QY 361 AGCATTCAGAGGTGACTGGTGGTGGCTGTTAAAGGCA 396
Db 413 AGCATTCAGAGGTGACTGGTGGTGGCTGTTAAAGGCA 448
```

```
RESULT 2
US-09-997-610-5
; Sequence 5, Application US/09997610
; Patent No. US20020156244A1
; GENERAL INFORMATION:
; APPLICANT: Fox, Brian
; APPLICANT: Holloway, James L.
; TITLE OF INVENTION: ADIPOCYTE COMPLIMENT RELATED PROTEIN
; FILE REFERENCE: 00-96
; CURRENT APPLICATION NUMBER: US/09/997,610
; CURRENT FILING DATE: 2001-11-29
; PRIOR APPLICATION NUMBER: US 60/253,924
; PRIOR FILING DATE: 2000-11-29
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 5
; LENGTH: 1731
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1731)
US-09-997-610-5
```

Query Match Best Local Similarity 100.0%; Score 396; DB 9; Length 1731;

Matches 396; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 GCTGACCTCCAGACACCCAGGCCCCAGAGAAAGTGGGGCTCTGTCACCAAGT 60
Db 406 GCTGACCTCCAGACACCCAGGCCCCAGAGAAAGTGGGGCTCTGTCACCAAGT 465
QY 61 TTACCAACATATACAGAGAAATAGTGAATGACAAATGCCCTGCTGATATAGAA 120
Db 466 TTACCAACATATACAGAGAAATAGTGAATGACAAATGCCCTGCTGATATAGAA 525
QY 121 AGGTGACCTTTACGTGTAAGTCACTGGAATCTCTCTCTTCAAGCCCATCATC 180
Db 526 AGGTGACCTTTACGTGTAAGTCACTGGAATCTCTCTCTTCAAGCCCATCATC 585
QY 181 TTACACAGGGTCTCTGTACATGCCCCAGAGGATTTAAGAGGCCATGGAGTCTTGT 240
Db 586 TTACACAGGGTCTCTGTACATGCCCCAGAGGATTTAAGAGGCCATGGAGTCTTGT 645
QY 241 TGCAGGGTCTGGGAATTAATCTACTCCAGCTTTGATGTTGAGTCATCAATGCAAGTG 300
Db 646 TGCAGGGTCTGGGAATTAATCTACTCCAGCTTTGATGTTGAGTCATCAATGCAAGTG 705
QY 301 AATATTGGCTAATGAGAGCAAAATTTGGCTAATTAAGAGCAAAATTTCTAAGCAGCA 360
Db 706 AATATTGGCTAATGAGAGCAAAATTTGGCTAATTAAGAGCAAAATTTCTAAGCAGCA 765
QY 361 AGCATTCAGAGGTGACTGGTGGTGGCTGTTAAAGGCA 396
Db 766 AGCATTCAGAGGTGACTGGTGGTGGCTGTTAAAGGCA 801
```

```
RESULT 3
US-09-997-610-3
; Sequence 3, Application US/09997610
; Patent No. US20020156244A1
; GENERAL INFORMATION:
; APPLICANT: Fox, Brian
; APPLICANT: Holloway, James L.
; TITLE OF INVENTION: ADIPOCYTE COMPLIMENT RELATED PROTEIN
; FILE REFERENCE: 00-96
; CURRENT APPLICATION NUMBER: US/09/997,610
; CURRENT FILING DATE: 2001-11-29
; PRIOR APPLICATION NUMBER: US 60/253,924
; PRIOR FILING DATE: 2000-11-29
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 3
; LENGTH: 1377
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Degenerate polynucleotide encoding a polypeptide
; NAME/KEY: variation
; LOCATION: (1)...(1377)
; OTHER INFORMATION: Each n is independently A, T, G, or C.
; NAME/KEY: misc_feature
; LOCATION: (1)...(1377)
; OTHER INFORMATION: n = A,T,C or G
US-09-997-610-3
```

Query Match Best Local Similarity 72.7%; Score 288; DB 9; Length 1377;

Matches 239; Conservative 87; Mismatches 69; Indels 0; Gaps 0;

```
QY 1 GCTGACCTCCAGACACCCAGGCCCCAGAGAAAGTGGGGCTCTGTCACCAAGT 60
Db 52 GTCGACCTCCAGACACCCAGGCCCCAGAGAAAGTGGGGCTCTGTCACCAAGT 111
QY 61 TTACCAACATATACAGAGAAATAGTGAATGACAAATGCCCTGCTGATATAGAA 120
Db 112 TTACCAACATATACAGAGAAATAGTGAATGACAAATGCCCTGCTGATATAGAA 171
QY 121 AGGTGACCTTTACGTGTAAGTCACTGGAATCTCTCTCTTCAAGCCCATCATC 180
Db 172 AGGTGACCTTTACGTGTAAGTCACTGGAATCTCTCTCTTCAAGCCCATCATC 231
QY 181 TTACACAGGGTCTCTGTACATGCCCCAGAGGATTTAAGAGGCCATGGAGTCTTGT 240
Db 232 TTACACAGGGTCTCTGTACATGCCCCAGAGGATTTAAGAGGCCATGGAGTCTTGT 291
QY 241 TGCAGGGTCTGGGAATTAATCTACTCCAGCTTTGATGTTGAGTCATCAATGCAAGTG 300
Db 292 TGCAGGGTCTGGGAATTAATCTACTCCAGCTTTGATGTTGAGTCATCAATGCAAGTG 351
QY 301 AATATTGGCTAATGAGAGCAAAATTTGGCTAATTAAGAGCAAAATTTCTAAGCAGCA 360
Db 352 AATATTGGCTAATGAGAGCAAAATTTGGCTAATTAAGAGCAAAATTTCTAAGCAGCA 411
QY 361 AGCATTCAGAGGTGACTGGTGGTGGCTGTTAAAGGCA 395
Db 412 AGCATTCAGAGGTGACTGGTGGTGGCTGTTAAAGGCA 446
```

```
RESULT 4
US-09-997-610-7
; Sequence 7, Application US/09997610
; Patent No. US20020156244A1
; GENERAL INFORMATION:
; APPLICANT: Fox, Brian
; APPLICANT: Holloway, James L.
; TITLE OF INVENTION: ADIPOCYTE COMPLIMENT RELATED PROTEIN
```

```

1  TITLE OR INVENTION:      zacrpl3
2  FILE REFERENCE: 00-96
3  CURRENT APPLICATION NUMBER: US/09/997,610
4  CURRENT FILING DATE: 2001-11-29
5  PRIOR APPLICATION NUMBER: US 60/253,924
6  PRIOR FILING DATE: 2000-11-29
7  NUMBER OF SEQ ID NOS: 7
8  SOFTWARE: FastSeq for Windows Version 3.0
9  SEQ ID NO 7
10 LENGTH: 1731
11 TYPE: DNA
12 ORGANISM: Artificial Sequence
13 FEATURE:
14 OTHER INFORMATION: Degenerate polynucleotide sequence of zacrpl3/zhnp1
15 OTHER INFORMATION: of SEQ ID NO:6
16 NAME/KEY: misc_feature
17 LOCATION: (1)..(1731)
18 OTHER INFORMATION: n = A,T,C or G
19 US-09-997-610-7

```

Query Match	72.7%;	Score 288;	DB 9;	Length 1731;
Best Local Similarity	60.5%;	Pred. NO. 2.6e-80;		
Matches 239;	Conservative 87;	Mismatches 69;	Indels 0;	Gaps 0

OY	1	GCTGACCTCCAGCACCACCGGCCCAAGAATGGGGGCCTCGTGTCACCAAGT	60
Db	406	GCNGNCCNCNCNCAYCCMGCCNCCNGARARGTNGSNCCNCGNGCCNCGN	465
OY	61	TTACCAAAATATACAGGAGAAATAAGTAAGTAACAATATGCCCTCTCGATATAGAA	120
Db	466	YTNCNCARTAYACNGNGGARATHMSNGARATGACNAARTGYCCNTGYCCNGAATHGAR	525
OY	121	AGGTACGCTTTACTGTGAAGCTAGTGGAAAACCTTCTCTTCCATTCAAGCCATCATC	180
Db	526	MGNMNSNGNTTYAACNGTNAARYTMWSNGNABARTYNNCCNTTYAARCCNATHATH	585
OY	181	TTTCAGAGGGTCCGTGCAATATGCCAGCGGATTTAAAGAGGCGCATGAGCTTTTGGT	240
Db	586	TTYTACNGSNGTNYNTAYTAATGCCARNGNSAYTTNABRGRCNATGGGNGTNTTGCN	645
OY	241	TGCAGGGTCCCTGGGAATTACTACTCCAGCTTTGATGTGGAGCTGCATCATTCAGCAAGTG	300
Db	646	TGYMNGTNCNCGNMAAYTATATWMSMNSNTTGAATGTCNABRYTNCAVATGYAANGTN	705
OY	301	AATATTGGCTAATGAGAGAACAAATTTTGGCTAATTAAGGAAGAAATTTCTAAGCAGCAA	360
Db	706	AAATHTGGTYTNATGMGMAARCACATHTYTNCCNAAYTAARGARATTHMSNAARCARCAR	765
OY	361	AGCATTCAGAGGTGACTTGGGTCTCTGTTAAAGC	395
Db	766	WSNATHCARGARGTNACNTGGTNTYTNTNANRCC	800

```

: RESULT 5
: US-09-960-352-15057
: Sequence 15057, Application us/09960352
: Patent No. US20020137139A1
: GENERAL INFORMATION:
: APPLICANT: Warren, Wesley C.
: APPLICANT: Tao, Nengding
: APPLICANT: Byatt, John C.
: APPLICANT: Mahalingam, Nagappan
: TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
: FILE REFERENCE: 16511.006/37-21(10298)C
: CURRENT APPLICATION NUMBER: US/09/960,352
: CURRENT FILING DATE: 2001-09-24
: NUMBER OF SEQ ID NOS: 15112
: SEQ ID NO 15057
: LENGTH: 425
: TYPE: DNA
: ORGANISM: Bos taurus
: OTHER INFORMATION: Clone ID: 64-LIB34-036-Q1-E1-H8

```

US-09-960-352-15057

Query Match	29.0%;	Score 114.8;	DB 10;	Length 425;
Best Local Similarity	72.3%;	Pred. NO. 2.5e-26;		
Matches 149;	Conservative	0;	Mismatches 57;	Indels 0;
			Gaps	0;

Qy	123	GTGAGCCCTTACTGTGGAAGCCACATGGAAAACCTTCCTCTTCCTTTCAAGGCCATCATCTT	183
Db	45	GTCCGCTTCACTGTGAAGCTCAATGTGCCATTTGCCCCCTTCAAAAGCCGTGGCCCTT	104
Qy	183	CACAGGGGTCCCTGTACATGCGCCAGAGGATTTTAAAGAGGCCATGGGAGTCTTGGCTTG	242
Db	105	CACATAGTCTCTGTACATATGCCAGAGATACTTACAGAGATGACACTGGGGTCCATCATG	164
Qy	243	CAGGATGCTGGGAATTACTCCAGCTTTGATGTTAGCTGCATCATTTGCAAGGTGAA	302
Db	165	CAGGATGCTCAAGAAATTAGCATTTCTCTTGATGTGGATCTCCATCATGCAATGTGAC	224
Qy	303	TATTTGGCTAATGAGGAAGCAAAATTT	328
Db	225	TGTTCACTGATGAGGCAAAAAGTT	250

RESULT 6
 US-09-960-352-9137
 : Sequence 9137, Application US/09960352
 : Patent No. US200201377139A1
 : GENERAL INFORMATION:
 : APPLICANT: Warren, Wesley C.
 : APPLICANT: Tao, Nengbing
 : APPLICANT: Byatt, John C.
 : APPLICANT: Mathialagan, Nagappan
 : TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
 : FILE REFERENCE: 16511.006/37-21.102981C
 : CURRENT APPLICATION NUMBER: US/09/960.352
 : CURRENT FILING DATE: 2001-09-24
 : NUMBER OF SEQ ID NOS: 15112
 : SEQ ID NO 9137
 : LENGTH: 415
 : TYPE: DNA
 : ORGANISM: Bos taurus
 : OTHER INFORMATION: Clone ID: 39-LIB34-018-Q1-E1-B12
 : US-09-960-352-9137

Query Match	23.58;	Score 93;	DB 10;	Length 415;
Best Local Similarity	63.28;	Pred. No. 1.7e-19;		
Matches 160;	Conservative 0;	Mismatches 90;	Indels 3;	Gaps 1;

QY	2	CTTACCTCCAGACACACCCCGCCCGCCCAAGAAAGTGGGCGCTCCGTGACACCCAGTT	61
Dd	163	CGGGCCACCCAGAGACACCCAGGGCGCTCCAGGTATTAAGAGGCGCTCCAGGTATTAAGAGGA	222
QY	62	TACCACATATACAGGA---GAATTAAGTGAATGACAAATGCGCCCTGTCTGATATAG	118
Dd	223	TACCAGTTTCCAGGTCCGCGCGGAACTCCAGACCAAGTGTAAATGGCCCATGGACA	282
QY	119	AAAGTAGCCTTTTACGTGTGAAGCTCAGTGGAAAATTCCCTCTCTTTGAACCCCATCA	178
Dd	283	GACAGTCCGCTTACGTGTGAAGCTCAGTGGCCAGTTGCCCTTCCCTTCAAAAGCGTGTGC	342
QY	179	TCTTTCACAGGGGCTCTGTACAAATGCCAGAGGGATTTAAAGAGGCCATGGAGCTTTTG	238
Dd	343	CTTTTCACAGAGGTCTCTGTACAAATGCCAGAGAGCTTACAGAGAGCACTGGGGTCTTCA	402
QY	239	CTTTCAGAGGTGCC	251
Dd	403	CATGCAAGGTGGC	415

RESULT 7
US-09-960-352-220
; Sequence 220, Application US/09960352

APPLICANT: Mathialagan, Nagappan
TITLE OF INVENTION: MUSCLE ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
FILE REFERENCE: 16511.006/37-21(10298)C
CURRENT FILING DATE: 2001-09-24
CURRENT APPLICATION NUMBER: US/09/960,352
NUMBER OF SEQ ID NOS: 15112
SEQ ID NO 11516
LENGTH: 374
TYPE: DNA
ORGANISM: Bos taurus
OTHER INFORMATION: Clone ID: 49-LIB34-046-Q1-E1-E2
US-09-960-352-11516

Query Match 19.2% Score 76; DB 10; Length 374;
Best Local Similarity 60.8%; Pred. No. 3,4e-14;
Matches 124; Conservative 0; Mismatches 80; Indels 0; Gaps 0;

QY 123 GTGAGCCTTACTGTGAGAGCTGAGGAAACCTCTCTTTCAGCCCATCTCTT 182
DB 171 GTCTGCTTTTCGCTCAACCTGAGTGGGCTTTCCAGAGACCTCCACCATTTGCTT 230
QY 183 CACAGGGCTCTGTACATGCCAGAGGATTTAAAGAGCCCATGGAGTCTTG 242
DB 231 CCAGGAAGCTGTACACATCAGGCCCATCTCGACCCGCTGCTGTCAACTG 290
QY 243 CAGGGCTCTGGGAATTAATCTACCTCAGCTTGTGAGTGCATCATTTGCAAGTGA 302
DB 291 CAGCGTCCCTGTGTGTACACTTTGGCTTTGACATTTGTTTCAAGTGTCTCA 350
QY 303 TATTTGGCTAATGAGAGCAAT 326
DB 351 AGGTGTCTAATGCGGAATGAAT 374

RESULT 11
US-09-960-352-3684
Sequence 3684, Application US/09960352
Patent No. US20020137139A1
GENERAL INFORMATION:
APPLICANT: Warren, Wesley C.
APPLICANT: Tao, Nengbing
APPLICANT: Byatt, John C.
APPLICANT: Mathialagan, Nagappan
TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
FILE REFERENCE: 16511.006/37-21(10298)C
CURRENT FILING DATE: 2001-09-24
CURRENT APPLICATION NUMBER: US/09/960,352
NUMBER OF SEQ ID NOS: 15112
SEQ ID NO 3684
LENGTH: 447
TYPE: DNA
ORGANISM: Bos taurus
OTHER INFORMATION: Clone ID: 16-LIB34-024-Q1-E1-D7
US-09-960-352-3684

Query Match 18.5% Score 73.2; DB 10; Length 447;
Best Local Similarity 60.6%; Pred. No. 2.8e-13;
Matches 120; Conservative 0; Mismatches 78; Indels 0; Gaps 0;

QY 123 GTGAGCCTTACTGTGAGAGCTGAGGAAACCTCTCTTTCAGCCCATCTCTT 182
DB 160 GTCTGCTTTTCGCTCAACCTGAGTGGGCTTTCCAGAGACCTCCACCATTTGCTT 219
QY 183 CACAGGGCTCTGTACATGCCAGAGGATTTAAAGAGCCCATGGAGTCTTG 242
DB 220 CCAGGAAGTGTGTACACATCAGGCCCATCTCGACCCGCTGCTGTCAACTG 279
QY 243 CAGGGCTCTGGGAATTAATCTACCTCAGCTTGTGAGTGCATCATTTGCAAGTGA 302
DB 280 CAGCGTCCCTGTGTGTACACTTTGGCTTTGACATTTGTTTCAAGTGTCTCA 339

QY 303 TATTTGGCTAATGAGGAA 320
DB 340 GGTGGCTCTAATTTGCGAA 357

RESULT 12
US-09-960-352-4241
Sequence 4241, Application US/09960352
Patent No. US20020137139A1
GENERAL INFORMATION:
APPLICANT: Warren, Wesley C.
APPLICANT: Tao, Nengbing
APPLICANT: Byatt, John C.
APPLICANT: Mathialagan, Nagappan
TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
FILE REFERENCE: 16511.006/37-21(10298)C
CURRENT FILING DATE: 2001-09-24
CURRENT APPLICATION NUMBER: US/09/960,352
NUMBER OF SEQ ID NOS: 15112
SEQ ID NO 4241
LENGTH: 467
TYPE: DNA
ORGANISM: Bos taurus
FEATURE:
NAME/KEY: unsure
LOCATION: (377)
OTHER INFORMATION: unsure at all n locations
OTHER INFORMATION: Clone ID: 18-LIB34-077-Q1-E1-E5
US-09-960-352-4241

Query Match 14.8% Score 58.8; DB 10; Length 467;
Best Local Similarity 54.9%; Pred. No. 9.3e-09;
Matches 139; Conservative 0; Mismatches 108; Indels 6; Gaps 1;

QY 2 CTGAGCTCCAGCACACCCAGGCCCCAGAAAGTGGGCTCTGTCACAGT 61
DB 217 CAGGCCCTCCGGCCCTCTCTGCGCTCTGCGCTCCAGAGGCCCCAGGCCCTCCAGANT 276
QY 62 TACCACATATACAGAGAAATTAAGTAATGACAAATGCCCTCTGATATAGAAA 121
DB 277 TGGGAGGACCACTGGGATTCGGGAGACATTTGAGATTGCTCTGCCCTAA- 332
QY 122 GGTACCTTTACTGTGAGAGCTCAGTGAAATCTCTCTCTTCAAGCCCATCATCT 181
DB 333 -TCTGCTTTGAGTGAAGATGAATGATCCCTCCAGCCCTCCAGCCCATTTGCT 390
QY 182 TCACAGGGTCTGTACATAGCCAGAGGATTTAAAGAGAGGAGTCTTGCTT 241
DB 391 TCAAGGGAACCTGCAATATGACAGGACACCTTCAATCTTACAGGGAGTGTACCT 450
QY 242 GCAGGGTGGCTGG 254
DB 451 GCACATCCCTGG 463

RESULT 13
US-09-864-761-25080
Sequence 25080, Application US/09864761
Patent No. US20020048763A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharon G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
APPLICANT: Chen, Wensheng
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
FILE REFERENCE: Aeonica-X-1
CURRENT APPLICATION NUMBER: US/09/864,761
CURRENT FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: US 60/180,312
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207,456

```

PRIORITY FILLING DATE: 2000-05-26
PRIORITY APPLICATION NUMBER: US 09/632.3,366
PRIORITY FILLING DATE: 2000-08-03
PRIORITY APPLICATION NUMBER: GB 24263.6
PRIORITY FILLING DATE: 2000-10-04
PRIORITY APPLICATION NUMBER: US 60/236,359
PRIORITY FILLING DATE: 2000-09-27
PRIORITY APPLICATION NUMBER: PCT/US01/00666
PRIORITY FILLING DATE: 2001-01-30
PRIORITY APPLICATION NUMBER: PCT/US01/00667
PRIORITY FILLING DATE: 2001-01-30
PRIORITY APPLICATION NUMBER: PCT/US01/00664
PRIORITY FILLING DATE: 2001-01-30
PRIORITY APPLICATION NUMBER: PCT/US01/00669
PRIORITY FILLING DATE: 2001-01-30
PRIORITY APPLICATION NUMBER: PCT/US01/00665
PRIORITY FILLING DATE: 2001-01-30
PRIORITY APPLICATION NUMBER: PCT/US01/00668
PRIORITY FILLING DATE: 2001-01-30
PRIORITY APPLICATION NUMBER: PCT/US01/00663
PRIORITY FILLING DATE: 2001-01-30
PRIORITY APPLICATION NUMBER: PCT/US01/00662
PRIORITY FILLING DATE: 2001-01-30
PRIORITY APPLICATION NUMBER: PCT/US01/00661
PRIORITY FILLING DATE: 2001-01-30
PRIORITY APPLICATION NUMBER: PCT/US01/00670
PRIORITY FILLING DATE: 2001-01-30
PRIORITY APPLICATION NUMBER: US 60/234,687
PRIORITY FILLING DATE: 2000-09-21
PRIORITY APPLICATION NUMBER: US 09/608,408
PRIORITY FILLING DATE: 2000-06-30
PRIORITY APPLICATION NUMBER: US 09/774,203
PRIORITY FILLING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Annonmax Sequence Listing Engine vers. 1.1
SEQ ID NO 25080
LENGTH: 273

TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO ALI211767.3
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 2
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 5.7
OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 6.1
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 2
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.8
OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 5.5
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 2.5
OTHER INFORMATION: SWISSPROT HIT: P29539, EVALUATE 6.80e-01
OTHER INFORMATION: EST HUMAN HIT: BE156318.1, EVALUATE 3.00e-78
OTHER INFORMATION: NT HIT: ALI63210.2, EVALUATE 7.00e-75
US-09-864-761-25080

Query Match          14.7%; Score 58.4; DB 10; Length 273;
Best Local Similarity 74.0%; Pred. No. 9.4e-09;
Matches      74; Conservative    0; Mismatches 26; Indels    0; Gaps    0

QY      237 GGTGATATTTCGGCTAAAGCAGCAAAATTTTGCGCTAATAAGGAAGAATTTCTTAAGCA 356
        ||| |||| |  || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db       115 GGCGAACATTCGAACCTTGAGAGAGATGATTGGGTATCTGTGTGAAGAATAATTTCTTAACA 174

QY      357 GCAAAGCATTCAGAGGTGACTGGTGGCTGTCTTAAGCA 396
        ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||
Db       175 GCAAAGCATTCAGAGGTGACTGGCTGTCTTAAGCA 214

RESULT 14
US-09-864-761-8349
; Sequence 8349, Application US/09864761
; Patent No. US20020048763A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharron G.
APPLICANT: Rank, David R.
```

```

1. APPLICANT : Hanzel, David K.
2. APPLICANT : Chen, Wensheng
3. TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
4. TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
5. FILE REFERENCE: Aecmlca-X-1
6. CURRENT APPLICATION NUMBER: US/09/864,761
7. PRIOR FILING DATE: 2001-05-23
8. PRIOR APPLICATION NUMBER: US 60/180,312
9. PRIOR FILING DATE: 2000-02-04
10. PRIOR APPLICATION NUMBER: US 60/207,456
11. PRIOR FILING DATE: 2000-05-26
12. PRIOR APPLICATION NUMBER: US 09/632,366
13. PRIOR FILING DATE: 2000-08-03
14. PRIOR APPLICATION NUMBER: GB 24263,6
15. PRIOR FILING DATE: 2000-10-04
16. PRIOR APPLICATION NUMBER: US 60/236,359
17. PRIOR FILING DATE: 2000-09-27
18. PRIOR APPLICATION NUMBER: PCT/US01/00666
19. PRIOR FILING DATE: 2001-01-30
20. PRIOR APPLICATION NUMBER: PCT/US01/00667
21. PRIOR FILING DATE: 2001-01-30
22. PRIOR APPLICATION NUMBER: PCT/US01/00664
23. PRIOR FILING DATE: 2001-01-30
24. PRIOR APPLICATION NUMBER: PCT/US01/00669
25. PRIOR FILING DATE: 2001-01-30
26. PRIOR APPLICATION NUMBER: PCT/US01/00665
27. PRIOR FILING DATE: 2001-01-30
28. PRIOR APPLICATION NUMBER: PCT/US01/00668
29. PRIOR FILING DATE: 2001-01-30
30. PRIOR APPLICATION NUMBER: PCT/US01/00663
31. PRIOR FILING DATE: 2001-01-30
32. PRIOR APPLICATION NUMBER: PCT/US01/00662
33. PRIOR FILING DATE: 2001-01-30
34. PRIOR APPLICATION NUMBER: PCT/US01/00661
35. PRIOR FILING DATE: 2001-01-30
36. PRIOR APPLICATION NUMBER: PCT/US01/00670
37. PRIOR FILING DATE: 2001-01-30
38. PRIOR APPLICATION NUMBER: US 60/234,687
39. PRIOR FILING DATE: 2000-09-21
40. PRIOR APPLICATION NUMBER: US 09/608,408
41. PRIOR FILING DATE: 2000-06-30
42. PRIOR APPLICATION NUMBER: US 09/774,203
43. PRIOR FILING DATE: 2001-01-29
44. NUMBER OF SEQ ID NOS: 49117
45. SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
46. SEQ ID NO 8349
47. LENGTH: 519
48. TYPE: DNA
49. ORGANISM: Homo sapiens
50. FEATURE:
51. OTHER INFORMATION: MAP TO AL121767.3
52. OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 2
53. OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 5.7
54. OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 6.1
55. OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 2
56. OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.8
57. OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 5.5
58. OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 2.5
59. US-09-864-761-8349
60. Query Match 14.7%; Score 58.4; DB 10; Length 519;
61. Best Local Similarity 74.0%; Pred. No. 1.3e-08;
62. Matches 74; Conservative 0; Mismatches 26; Indels 0; Gaps 0;
63.
64.
65.
66.
67.
68.
69.
70.
71.
72.
73.
74.
75.
76.
77.
78.
79.
80.
81.
82.
83.
84.
85.
86.
87.
88.
89.
90.
91.
92.
93.
94.
95.
96.
97.
98.
99.
100.
101.
102.
103.
104.
105.
106.
107.
108.
109.
110.
111.
112.
113.
114.
115.
116.
117.
118.
119.
120.
121.
122.
123.
124.
125.
126.
127.
128.
129.
130.
131.
132.
133.
134.
135.
136.
137.
138.
139.
140.
141.
142.
143.
144.
145.
146.
147.
148.
149.
150.
151.
152.
153.
154.
155.
156.
157.
158.
159.
160.
161.
162.
163.
164.
165.
166.
167.
168.
169.
170.
171.
172.
173.
174.
175.
176.
177.
178.
179.
180.
181.
182.
183.
184.
185.
186.
187.
188.
189.
190.
191.
192.
193.
194.
195.
196.
197.
198.
199.
200.
201.
202.
203.
204.
205.
206.
207.
208.
209.
210.
211.
212.
213.
214.
215.
216.
217.
218.
219.
220.
221.
222.
223.
224.
225.
226.
227.
228.
229.
230.
231.
232.
233.
234.
235.
236.
237.
238.
239.
240.
241.
242.
243.
244.
245.
246.
247.
248.
249.
250.
251.
252.
253.
254.
255.
256.
257.
258.
259.
260.
261.
262.
263.
264.
265.
266.
267.
268.
269.
270.
271.
272.
273.
274.
275.
276.
277.
278.
279.
280.
281.
282.
283.
284.
285.
286.
287.
288.
289.
290.
291.
292.
293.
294.
295.
296.
297.
298.
299.
300.
301.
302.
303.
304.
305.
306.
307.
308.
309.
310.
311.
312.
313.
314.
315.
316.
317.
318.
319.
320.
321.
322.
323.
324.
325.
326.
327.
328.
329.
330.
331.
332.
333.
334.
335.
336.
337.
338.
339.
340.
341.
342.
343.
344.
345.
346.
347.
348.
349.
350.
351.
352.
353.
354.
355.
356.
357.
358.
359.
360.
361.
362.
363.
364.
365.
366.
367.
368.
369.
370.
371.
372.
373.
374.
375.
376.
377.
378.
379.
380.
381.
382.
383.
384.
385.
386.
387.
388.
389.
390.
391.
392.
393.
394.
395.
396.
397.
398.
399.
400.
401.
402.
403.
404.
405.
406.
407.
408.
409.
410.
411.
412.
413.
414.
415.
416.
417.
418.
419.
420.
421.
422.
423.
424.
425.
426.
427.
428.
429.
430.
431.
432.
433.
434.
435.
436.
437.
438.
439.
440.
441.
442.
443.
444.
445.
446.
447.
448.
449.
450.
451.
452.
453.
454.
455.
456.
457.
458.
459.
460.
461.
462.
463.
464.
465.
466.
467.
468.
469.
470.
471.
472.
473.
474.
475.
476.
477.
478.
479.
480.
481.
482.
483.
484.
485.
486.
487.
488.
489.
490.
491.
492.
493.
494.
495.
496.
497.
498.
499.
500.
501.
502.
503.
504.
505.
506.
507.
508.
509.
510.
511.
512.
513.
514.
515.
516.
517.
518.
519.
520.
521.
522.
523.
524.
525.
526.
527.
528.
529.
530.
531.
532.
533.
534.
535.
536.
537.
538.
539.
540.
541.
542.
543.
544.
545.
546.
547.
548.
549.
550.
551.
552.
553.
554.
555.
556.
557.
558.
559.
560.
561.
562.
563.
564.
565.
566.
567.
568.
569.
570.
571.
572.
573.
574.
575.
576.
577.
578.
579.
580.
581.
582.
583.
584.
585.
586.
587.
588.
589.
590.
591.
592.
593.
594.
595.
596.
597.
598.
599.
600.
601.
602.
603.
604.
605.
606.
607.
608.
609.
610.
611.
612.
613.
614.
615.
616.
617.
618.
619.
620.
621.
622.
623.
624.
625.
626.
627.
628.
629.
630.
631.
632.
633.
634.
635.
636.
637.
638.
639.
640.
641.
642.
643.
644.

```



```

RESULT 15
US-09-864-761-2925
Sequence 2925 Application US/09864761
Patent No. US20020048763A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharon G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
APPLICANT: Chen, Wensheng
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
FILE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
TITLE REFERENCE: Aecmca-x-1
CURRENT FILING DATE: 2001-05-23
PRIORITY FILING DATE: 2001-05-23
PRIORITY FILING DATE: US 60/180,312
PRIORITY FILING DATE: US 60/207,456
PRIORITY FILING DATE: 2000-05-26
PRIORITY FILING DATE: US 09/632,366
PRIORITY FILING DATE: 2000-08-03
PRIORITY APPLICATION NUMBER: GB 24263.6
PRIORITY FILING DATE: 2000-10-04
PRIORITY APPLICATION NUMBER: US 60/236,359
PRIORITY FILING DATE: 2000-09-27
PRIORITY APPLICATION NUMBER: PCT/US01/00666
PRIORITY FILING DATE: 2001-01-30
PRIORITY APPLICATION NUMBER: PCT/US01/00667
PRIORITY FILING DATE: 2001-01-30
PRIORITY APPLICATION NUMBER: PCT/US01/00664
PRIORITY FILING DATE: 2001-01-30
PRIORITY APPLICATION NUMBER: PCT/US01/00669
PRIORITY FILING DATE: 2001-01-30
PRIORITY APPLICATION NUMBER: PCT/US01/00665
PRIORITY FILING DATE: 2001-01-30
PRIORITY APPLICATION NUMBER: PCT/US01/00668
PRIORITY FILING DATE: 2001-01-30
PRIORITY APPLICATION NUMBER: PCT/US01/00663
PRIORITY FILING DATE: 2001-01-30
PRIORITY APPLICATION NUMBER: PCT/US01/00662
PRIORITY FILING DATE: 2001-01-30
PRIORITY APPLICATION NUMBER: PCT/US01/00661
PRIORITY FILING DATE: 2001-01-30
PRIORITY APPLICATION NUMBER: PCT/US01/00670
PRIORITY FILING DATE: 2001-01-30
PRIORITY APPLICATION NUMBER: US 60/234,687
PRIORITY FILING DATE: 2000-09-21
PRIORITY APPLICATION NUMBER: US 09/608,408
PRIORITY FILING DATE: 2000-06-30
PRIORITY APPLICATION NUMBER: US 09/774,203
PRIORITY FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
SEQ ID NO 2925
LENGTH: 1946
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AC006337.2
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 20
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 14
OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 15
OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 13
OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 12
OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 12
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 10
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 14
OTHER INFORMATION: EXPRESSED IN LONG, SIGNAL = 13
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 13
US-09-864-761-2925

```

```

QY 300 GAAATTTGGCTAATGAGGAAGCAAAATTTTGGCTAATAAGAAAGAAATTTCTAAGCAGCA 359
Db 174 GAAATTTGAACTTTCAAGAGATGATTTAGGGTAATCTGCTGGAGAAATTTCTAAGCAGCA 233
QY 360 AAGCATTCACAGAGTGACTTTGGGCTGTTTAAAGCA 396
Db 234 AAGCATTCACAGAGTGACTTGGTACTGTTAAAGCA 270

```

Search completed: February 19, 2003, 23:02:47
Job time : 39.8656 secs

Search completed: February 19, 2003, 23:02:47
Job time : 39.8656 secs

Query Match	14.4%;	Score 57;	DB 10;	Length 1946;
Best Local Similarity	74.2%;	Pred. No. 7.3e-08;		
Matches	72;	Conservative	0;	Mismatches 25;
			Indels	0;
			Gaps	0;

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 19, 2003, 16:09:30 : Search time 26.1422 Seconds
(without alignments)
4645.518 Million cell updates/sec

Title: US-09-997-610-1_COPY_53_448

Perfect score: 396

Sequence: 1 gctgacctccagcacacc.....cttggtgtgttaagca 396

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 15338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database : Issued_Patents_NA:*

1: /cgn2_6/ptodata/2/lna/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/lna/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/lna/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/lna/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/lna/PTUS_COMB.seq:*
6: /cgn2_6/ptodata/2/lna/Backfile1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	46.6	11.8	1313	2	US-08-463-911-6
2	46.6	11.8	4517	4	US-09-140-804-9
3	42.4	10.7	1107	3	US-09-188-930-217
4	40	10.1	1276	2	US-08-463-911-1
5	37.4	9.4	728	4	US-09-336-536-2
6	37.4	9.4	1338	4	US-09-336-536-1
7	37.4	9.4	1347	4	US-09-140-804-1
8	36.2	9.1	729	4	US-09-140-804-10
9	34.4	8.7	14855	2	US-08-687-080-59
10	34.4	8.7	44453	4	US-09-146-053-5
11	34.4	8.7	49136	4	US-09-422-869-1
12	34	8.6	459	4	US-09-415-551-2
13	34	8.6	486	4	US-09-415-551-1
14	33.2	8.4	1881	4	US-09-029-348-20
15	33.2	8.4	168575	4	US-09-426-290-1
16	31.4	7.9	1001	3	US-09-188-930-218
17	31.4	7.9	1015	3	US-09-188-930-30
18	31.4	7.9	35081	2	US-08-752-760A-1
19	31	7.8	1215	2	US-09-092-770-8
20	31	7.8	1215	2	US-09-222-851-8
21	30.6	7.7	2712	1	US-08-346-455B-37
22	30.6	7.7	2712	1	US-08-977-221-37
23	30.6	7.7	2712	4	US-09-483-831B-37
24	30.6	7.7	2712	5	PCT-US95-06613-37
25	30.6	7.7	2946	1	US-08-346-455B-35
26	30.6	7.7	2946	3	US-08-977-221-35
27	30.6	7.7	2946	4	US-09-483-831B-35

C 28	30.6	7.7	2946	5	PCT-US95-06613-35	Sequence 35, Appl
C 29	30.6	7.7	3251	1	US-08-346-455B-68	Sequence 68, Appl
C 30	30.6	7.7	3251	3	US-08-977-221-68	Sequence 68, Appl
C 31	30.6	7.7	3251	4	US-09-483-831B-68	Sequence 68, Appl
C 32	30.6	7.7	3251	5	PCT-US95-06613-68	Sequence 68, Appl
C 33	30.4	7.7	420	2	US-08-557-892-5	Sequence 5, Appl
C 34	30.4	7.7	420	2	US-08-387-858A-5	Sequence 5, Appl
C 35	30.4	7.7	420	4	US-09-294-384B-5	Sequence 5, Appl
C 36	30.4	7.7	420	4	US-08-717-079-5	Sequence 5, Appl
C 37	30.4	7.7	519	1	US-08-318-905-18	Sequence 18, Appl
C 38	30.4	7.7	519	1	US-08-483-232-18	Sequence 18, Appl
C 39	30.4	7.7	519	1	US-08-483-140-18	Sequence 18, Appl
C 40	30.4	7.7	519	2	US-08-485-938A-18	Sequence 18, Appl
C 41	30.4	7.7	519	2	US-08-910-041-18	Sequence 18, Appl
C 42	30.4	7.7	519	3	US-09-328-474-18	Sequence 18, Appl
C 43	30.4	7.7	519	3	US-09-100-546-18	Sequence 18, Appl
C 44	30.4	7.7	519	3	US-09-100-546-18	Sequence 18, Appl
C 45	30.4	7.7	519	4	US-09-577-758-18	Sequence 18, Appl

ALIGNMENTS

RESULT 1
US-08-463-911-6
; Sequence 6, Application US/08463911
; Patent No. 5869330
; GENERAL INFORMATION:
; APPLICANT: Scherer, Philipp E.
; APPLICANT: Lodish, Harvey F.
; TITLE OF INVENTION: A NOVEL SERUM PROTEIN PRODUCED
; TITLE OF INVENTION: EXCLUSIVELY IN ADIPOCYTES
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
; STREET: Two Millitia drive
; CITY: Lexington
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02173
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/463,911
; FILING DATE:
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Granahan, Patricia
; REGISTRATION NUMBER: 32,227
; REFERENCE/DOCKET NUMBER: WH195-05
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 861-6240
; TELEFAX: (617) 861-9540
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1313 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: Linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 73..804
; US-08-463-911-6
Query Match 11.8%; Score 46.6; DB 2; Length 1313;
Best Local Similarity 48.9%; Pred. No. 1.8e-05;
Matches 158; Conservative 0; Mismatches 159; Indels 6; Gaps 1;
QY 3 TGGACCTCCAGCACACCCAGGCCCCAGAAAGATGGGGCTTCGTGCACACGCTTT 62


```

:      REGISTRATION NUMBER: 32.227
:      REFERENCE/DOCKET NUMBER: WH195-05
:      TELECOMMUNICATION INFORMATION:
:      TELEPHONE: (617) 861-6240
:      TELEFAX: (617) 861-9540
:      INFORMATION FOR SEQ ID NO: 1:
:      SEQUENCE CHARACTERISTICS:
:      LENGTH: 1276 base pairs
:      TYPE: nucleic acid
:      STRANDEDNESS: single
:      TOPOLOGY: linear
:      MOLECULE TYPE: DNA (genomic)
:      FEATURE:
:      NAME/KEY: CDS
:      LOCATION: 46..786
:      US-08-463-911-1

```

[illegible]

```

? RESULT 5
? US-09-336-536-2
? Sequence 2, Application US/09336536
? Patent No. 640684
? GENERAL INFORMATION:
? APPLICANT: Leiby, K.
? APPLICANT: McKay, C.
? APPLICANT: Bossone, S.
? TITLE OF INVENTION: SECRETED PROTEINS AND USES THEREOF
? FILE REFERENCE: 7853-144
? CURRENT APPLICATION NUMBER: US/09/336,536
? CURRENT FILING DATE: 1999-06-18
? NUMBER OF SEQ ID NOS: 75
? SOFTWARE: PatentIn Ver. 2.0
? SEQ ID NO 2
? LENGTH: 728
? TYPE: DNA
? ORGANISM: Homo sapiens
? US-09-336-536-2

```

	Query Match	9.48;	Score 37.4;	D _B 4;	Length 728;
	Best Local Similarity	51.5%;	Pred. No. 0.013;		
	Matches	86;	Conservative	0;	Mismatches 81; Indels 0; Gaps 0;
OY	154	CTTCCCTCTTCCTTTTCAACGCCATCATCTTCTACAGGGGCTTTACAATCCCGAGGGAT	213		
Db	345	CCTCCGCCGTGTGAACGCACCCTTGCCCTTCGACGCCGCTGTGTACACAGAGGACACT	404		
OY	214	TTTAAGGAGGCCAATGGAGTCTTTTCTTTCGACAGGTGCTCGGGAATTACTTCACGTTT	273		
Db	405	TACGCGCGCCTTCACCGGGAAGTTCACCTGCGCAGGTGCTCGGGGTCTACTACTTCGCCGTC	464		
OY	274	GATGTGTAGCTGCATCATTTGCAAGGTAAATTTGGCTAACAGAGA	320		
Db	465	CATGCCACCGTCTTACCGGGCCAGCTGGAAGTTTGTGTGTAATAA	511		

```

RESULT 6
US-09-336-536-1
: Sequence 1, Application US/09336536
: Patent No. 6406884
: GENERAL INFORMATION:
: APPLICANT: Leidy, K.
: APPLICANT: McKay, C.
: APPLICANT: Bossone, S.
: TITLE OF INVENTION: SECRETED PROTEINS AND USES THEREOF
: FILE REFERENCE: 7853-144
: CURRENT APPLICATION NUMBER: US/09/336,536
: CURRENT FILING DATE: 1999-06-18
: NUMBER OF SEQ ID NOS: 75
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 1
: LENGTH: 1338
: TYPE: DNA
: ORGANISM: Homo sapiens
US-09-336-536-1

```

[illegible]

```

RESULT 7
US-09-140-804-1
: Sequence 1, Application US/09140804
: Patent No. 6197930
: GENERAL INFORMATION:
: APPLICANT: Sheppard, Paul O.
: APPLICANT: Humes, Jacqueline M.
: TITLE OF INVENTION: ADIPOCYTE-SPECIFIC PROTEIN HOMOLOGS
: FILE REFERENCE: 97-49
: CURRENT APPLICATION NUMBER: US/09/140, 804
: CURRENT FILING DATE: 1998-08-26
: EARLIER APPLICATION NUMBER: 60/056, 983
: EARLIER FILING DATE: 1997-08-26
: NUMBER OF SEQ ID NOS: 47
: SOFTWARE: FastSeq for Windows Version 3.0
: SEQ ID NO 1
: LENGTH: 1347
: TYPE: DNA
: ORGANISM: Homo sapiens
: FEATURE:
: NAME/KEY: CDS
: LOCATION: (198)...(926)
US-09-140-804-1

```

[illegible]

[illegible]

```
RESULT 15
US-09-426-290-1/c
; Sequence 1, Application US/09426290
; Patent No. 6410712
; GENERAL INFORMATION:
; APPLICANT: Berglund Ran Olofsdottr
; TITLE OF INVENTION: HUMAN NARCOLEPSY GENE
; FILE REFERENCE: 2345.2001-000
; CURRENT APPLICATION NUMBER: US/09/426,290
; CURRENT FILING DATE: 1999-10-25
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 168575
; TYPE: DNA
; ORGANISM: Homo Sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (21181)...(21403)
; NAME/KEY: CDS
; LOCATION: (95252)...(95430)
; NAME/KEY: CDS
; LOCATION: (101753)...(101996)
; NAME/KEY: CDS
; LOCATION: (110324)...(110439)
; NAME/KEY: CDS
; LOCATION: (124058)...(124278)
; NAME/KEY: CDS
; LOCATION: (127009)...(127130)
; NAME/KEY: CDS
; LOCATION: (128910)...(129139)
US-09-426-290-1

Query Match      8.4%; Score 33.2; DB 4; Length 168575;
Best Local Similarity 75.9%; Pred.No.4.9;
Matches 41; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY 336 TAAGGAGAATTTCTTAAGCAGCAAGCATTCAGAGAGTGACTGGCTGCTTT 389
    || ||||| |||| ||||| ||||| || ||||| ||||| ||||| |||||
Db 64084 TATAGAGAAATTTTAAAGGAGCAAGCATTCATGATGCTGTGGCTACTTTT 64031

Search Completed: February 19, 2003, 23:02:25
Job time : 231.142 secs
```


GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 19, 2003, 15:08:07 : Search time 137.053 Seconds
(without alignments)
6506.903 Million cell updates/sec

Title: US-09-997-610-1_COPY_53_448

Perfect score: 396
Sequence: 1 gctgagcttcacacacaccc.....cttggtgtctgttaagca 396

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :
1: N.Geneseq_101002.*
2: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.*
3: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.*
4: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.*
5: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.*
6: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.*
7: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.*
8: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.*
9: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.*
10: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.*
11: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.*
12: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.*
13: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.*
14: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.*
15: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.*
16: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.*
17: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.*
18: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.*
19: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.*
20: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.*
21: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.*
22: /SID2/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.*
23: /SID2/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.*
24: /SID2/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.*
25: /SID2/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	351	88.6	1338	24	AA144066
2	58.6	14.8	404	21	AAC24351
3	58.4	14.7	273	22	ABA75076
4	58.4	14.7	273	22	ABA39760
5	58.4	14.7	273	22	AAK23591
6	58.4	14.7	273	22	AAK49719
7	58.4	14.7	273	22	AA126826
8	58.4	14.7	273	22	AA155610
9	58.4	14.7	519	22	ABA62555

10	58.4	14.7	519	22	ABA29883	Probe #8349 for ge
11	58.4	14.7	519	22	AAK10911	Human brain expres
12	58.4	14.7	519	22	AAK36769	Human bone marrow
13	58.4	14.7	519	22	AA117621	Probe #17554 for ge
14	58.4	14.7	519	22	AA142541	Probe #11227 used
15	57	14.4	662	22	AA115063	Human breast cancer
16	57	14.4	930	22	AA123916	Human breast cancer
17	57	14.4	920	23	AA581134	DNA encoding novel
18	57	14.4	999	23	AA566807	DNA encoding novel
19	57	14.4	1077	23	AA582559	DNA encoding novel
20	57	14.4	1946	22	ABA44227	Human breast cell
21	57	14.4	1946	22	ABA54677	Human foetal liver
22	57	14.4	1946	22	ABA24459	Probe #2925 for ge
23	57	14.4	1946	22	AAK02964	Human brain expres
24	57	14.4	1946	22	AAK28410	Human bone marrow
25	57	14.4	1946	22	AA112975	Probe #2908 for ge
26	57	14.4	1946	22	AA134334	Probe #3020 used t
27	57	14.4	1946	22	AA102893	Probe #2884 used t
28	57	14.4	1946	24	AA502918	Human genome-deriv
29	56.4	14.2	32127	22	AA199255	Human excretory re
30	56.4	14.2	32127	22	AA163605	Human kidney relat
31	56.4	14.2	72215	22	AAK68632	Human immune/haema
32	56	14.1	505	22	ABA60370	Human foetal liver
33	56	14.1	505	22	ABA28609	Probe #7075 for ge
34	56	14.1	505	22	AAK08648	Human brain expres
35	56	14.1	505	22	AAK34535	Human bone marrow
36	56	14.1	505	22	AA140252	Probe #8938 used t
37	56	14.1	505	24	AB509176	Human genome-deriv
38	56	14.1	1054	23	AA564468	DNA encoding novel
39	55.4	14.0	342	21	AAK26950	Human secreted pro
40	55.4	14.0	520	22	ABA26906	Human foetal liver
41	55.4	14.0	520	22	ABA30177	Probe #8643 for ge
42	55.4	14.0	520	22	AAK11311	Human brain expres
43	55.4	14.0	520	22	AAK37101	Human bone marrow
44	55.4	14.0	520	22	AA117942	Probe #7875 for ge
45	55.4	14.0	520	22	AA142928	Probe #11614 used

ALIGNMENTS

RESULT 1	AA144066	standard; cDNA; 1338 BP.
ID	AA144066;	
AC	AA144066;	
XX		
DT	27-SEP-2002	(first entry)
XX		
DE	Human genset metabolic gene (GMC-9) cDNA sequence.	
XX		
KW	Human; gene; ss; gene therapy; genset metabolic gene; GMC-7A; GMC-7B;	
KW	GMC-8; GMC-9; GMC-10; GMC-11; metabolic-related disorder; obesity;	
KW	Impaired glucose tolerance; insulin resistance; Syndrome X;	
KW	Type II diabetes; hyperlipidaemia; atherosclerosis; hypertension;	
KW	heart disease; cardiac insufficiency; coronary insufficiency;	
KW	high blood pressure; insulin sensitivity;	
KW	non-insulin dependent diabetes mellitus.	
XX		
OS	Homo sapiens.	
XX		
FT	key	Location/Qualifiers
FT	CDS	1..1338
FT		/tag- a
FT		/partial
FT		/product- "Human GMC-9 protein"
XX		/note- "No stop codon is given"
PN	WO200255694-A2.	
XX		
PD	18-JUL-2002.	
XX		
PF	15-JAN-2002; 2002WO-IB01215.	

```
XX PR 16-JAN-2001; 2001US-262235P.
XX PA (GEST ) GENSET.
XX PI Erickson MR, Bour BA, Bihain B, Tanaka H;
XX DR WPI: 2002-557821/59.
XX P-PSDB; AAO15423.
XX PT Treating or preventing a metabolic-related disease or disorder, e.g.
XX PT obesity, impaired glucose tolerance, insulin resistance, Syndrome X, or
XX PT Type II diabetes, comprises administering Genset Metabolic Genes -
XX PS Disclosure; Page 122-124; 128pp; English.
XX XX
XX CC The invention comprises the amino acid and coding sequences of six human
XX CC genset metabolic genes (GMG-7A, GMG-7B, GMG-8, GMG-9, GMG-10 and GMG-11).
XX CC The GMG DNA and protein sequences of the invention are useful for
XX CC treating or preventing metabolic-related disorders, such as: obesity;
XX CC impaired glucose tolerance; insulin resistance; Syndrome X; Type II
XX CC diabetes; hyperlipidaemia; atherosclerosis; hypertension; and heart
XX CC diseases (e.g. cardiac insufficiency, coronary insufficiency or high
XX CC blood pressure). The GMG DNA and protein sequences of the invention may
XX CC also be used as insulin sensitizers - for improving insulin sensitivity
XX CC in persons with non-insulin dependent diabetes mellitus. The present CDNA
XX CC sequence encodes the human GMG-9 protein.
XX SQ Sequence 1338 BP; 359 A; 288 C; 350 G; 341 T; 0 other;
XX
XX Query Match 88.6%; Score 351; DB 24; Length 1338;
XX Best Local Similarity 98.6%; Pred. No. 2.3e-104;
XX Matches 334; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
XX
XX QY 38 TGGGGCCCTCGTGTGACACAGTTTACACATATACAGAGAAATAGTAATGACAA 97
XX DB 50 TGGAAATGCTGTGTCGCCCAAGTTTACACATATACAGAGAAATAGTAATGACAA 109
XX
XX QY 98 AATGCCCTCTCTGATATAGAAAGTCAAGCTTTACTGTGAAGCTGAGGAAACTTC 157
XX DB 110 AATGCCCTCTCTGATATAGAAAGTCAAGCTTTACTGTGAAGCTGAGGAAACTTC 169
XX
XX QY 158 CTCTTCCTTCAAGCCCATATCTTCACAGGGTCCCTGATCAANTGCCACAGGATTAA 217
XX DB 170 CTCTTCCTTCAAGCCCATATCTTCACAGGGTCCCTGATCAANTGCCACAGGATTAA 229
XX
XX QY 218 AGGAGCCCATGAGAGTCTTGTGACAGGGTGCCTGGGAATTAATCAAGCTTGATG 277
XX DB 230 AGGAGGCGCATGGAGTCTTGTGACAGGGTGCCTGGGAATTAATCAAGCTTGATG 289
XX
XX QY 278 TTGAGCTGCATCATTCGCAAGGTGAATTTGGCTAATGAGGAACAATTTGGCTAATA 337
XX DB 290 TTGAGCTGCATCATTCGCAAGGTGAATTTGGCTAATGAGGAACAATTTGGCTAATA 349
XX
XX QY 338 AGGAGAAATTTCTTAACAGCAAAAGCATTCGAAGGTGACTGGCTGTTAAAGCA 396
XX DB 350 AGGAGAAATTTCTTAACAGCAAAAGCATTCGAAGGTGACTGGCTGTTAAAGCA 408
XX
XX RESULT 2
XX ID AAC24351 standard; cDNA; 404 BP.
XX AC AAC24351;
XX DT 06-OCT-2000 (first entry)
XX XX
XX DE Human secreted protein 5' EST, SEQ ID NO: 28426.
XX XX
XX KW Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
XX KW gene therapy; chromosome mapping; ss.
XX OS Homo sapiens.
```

```
XX XX
XX PN EP1033401-A2.
XX PD 06-SEP-2000.
XX XX
XX PF 21-FEB-2000; 2000EP-0200610.
XX XX
XX PR 26-FEB-1999; 99US-0122487.
XX XX
XX PA (GEST ) GENSET.
XX PI Dumas Milne Edwards J, Duclert A, Giordano J;
XX DR WPI: 2000-500381/45.
XX XX
XX XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for
XX PT obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
XX PT diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX PS Claim 1; SEQ ID 28426; 71pp + CD-ROM; English.
XX XX
XX XX The present sequence is one of a large number of 5' ESTs derived from
XX CC mRNAs encoding secreted proteins. No ORF has yet been conclusively
XX CC identified within the present sequence. The 5' ESTs were prepared from
XX CC total human RNAs or polyA+ RNAs derived from 30 different tissues. EST
XX CC sequences usually correspond mainly to the 3' untranslated region (UTR)
XX CC of the mRNA because they are often obtained from oligo-dT primed cDNA
XX CC libraries. Such ESTs are not well suited for isolating cDNA sequences
XX CC derived from the 5' ends of mRNAs and even in those cases where longer
XX CC cDNA sequences have been obtained, the full 5' UTR is rarely included.
XX CC 5' ESTs are derived from mRNAs with intact 5' ends and can therefore be
XX CC used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used
XX CC in diagnostic, forensic, gene therapy and chromosome mapping procedures.
XX CC They are used to obtain upstream regulatory sequences and to design
XX CC expression and secretion vectors.
XX SQ Sequence 404 BP; 137 A; 55 C; 101 G; 111 T; 0 other;
XX
XX Query Match 14.8%; Score 58.6; DB 21; Length 404;
XX Best Local Similarity 75.3%; Pred. No. 8e-09;
XX Matches 73; Conservative 0; Mismatches 24; Indels 0; Gaps 0;
XX
XX QY 300 GAATATTGGCTATGAGGAACAATTTGGCTAATTAAGGAACAATTTCTAAGCAGCA 359
XX DB 207 GAATCTTGAAGCTGAGGAACAATTTGAATGATAGTGTGGAAGAATTTCTAAGCAGCA 266
XX
XX QY 360 AAGCATTCAGAGGTGACTGGGTGCTGTTAAAGCA 396
XX DB 267 AAGCATTCAGAGGTGACTGGGTGCTGTTAAAGCA 303
XX
XX RESULT 3
XX ID ABA75076 standard; DNA; 273 BP.
XX AC ABA75076;
XX DT 01-FEB-2002 (first entry)
XX XX
XX DE Human foetal liver single exon nucleic acid probe #23381.
XX XX
XX KW Human; foetal liver; gene expression; single exon nucleic acid probe; ss.
XX KW Homo sapiens.
XX OS
XX PN WO200157277-A2.
XX PD 09-AUG-2001.
XX XX
XX XX 30-JAN-2001; 2001WO-0500669.
XX PF 04-FEB-2000; 2000US-0180312.
XX PR 26-MAY-2000; 2000US-0207456.
```

PR	30-JUN-2000;	2000US-0608408.
PR	03-AUG-2000;	2000US-0632366.
PR	21-SEP-2000;	2000US-0234687.
PR	27-SEP-2000;	2000US-0236359.
PR	04-OCT-2000;	2000GB-0024263.
XX		
XX	(MOLE-)	MOLECULAR DYNAMICS INC.
XX		
XX	Penn SG,	Hanzel DK, Chen W, Rank DR;
XX		
DR	WPI;	2001-483447/52.
XX		
PT	Human genome-derived single exon nucleic acid probes useful for	
PT	analyzing gene expression in human fetal liver.	
XX		
PS	Claim 4;	SEQ ID NO 23381; 639pp + sequence listing; English.
XX		
CC	The invention relates to a single exon nucleic acid probe for	
CC	measuring human gene expression in a sample derived from human foetal	
CC	liver. The single exon nucleic acid probes may be used for predicting,	
CC	measuring and displaying gene expression in samples derived from human	
CC	fetal liver. The present sequence is a single exon nucleic acid	
CC	probe of the invention.	
CC	Note: The sequence data for this patent did not form part of the	
CC	printed specification, but was obtained in electronic format directly	
CC	from WIPO at ftp.wipo.int/pub/published_pct_sequences.	
XX		
SQ	Sequence 273 BP; 80 A; 36 C; 81 G; 76 T; 0 other;	
	Query Match	14.7%; Score 58.4; DB 22; Length 273;
	Best Local Similarity	74.0%; Pred. NO. 7.6e-09;
	Matches 74; Conservative	0; Mismatches 26; Indels 0; Gaps 0;
OY	297	GGTGAATATTGGCTAATGAGGAACAATTTGGCTAATAGGAAGAATTTCTAAGCA 356
Db	115	GGGGACACTTGACTGACTGAGAGATGATTTGGGCTATCTGGTGGAAGAATTTCTAAGCA 174
OY	357	GCAAGCATTTCAGAGGTCGACTTGGGTCCTTTAAAGCA 396
Db	175	GCAAGCATTTCAGAGGTCGACTTGGGTCCTTTAAAGCA 214
RESULT 4		
ABA39760		
ID	ABA39760 standard; DNA; 273 BP.	
XX		
AC	ABA39760;	
XX		
DT	23-JAN-2002 (first entry)	
XX		
DE	Probe #18226 for gene expression analysis in human heart cell sample.	
XX		
KW	Human; gene expression; heart; microarray; vascular system; probe;	
KW	cardiovascular disease; hypertension; cardiac arrhythmia;	
XX	congenital heart disease; ss.	
XX		
OS	Homo sapiens.	
XX		
PN	WO200157274-A2.	
XX		
PD	09-AUG-2001.	
XX		
PF	30-JAN-2001; 2001WO-US00666.	
XX		
PR	04-FEB-2000; 2000US-0180312.	
PR	26-MAY-2000; 2000US-0207456.	
PR	30-JUN-2000; 2000US-0608408.	
PR	03-AUG-2000; 2000US-0632366.	
PR	21-SEP-2000; 2000US-0234687.	
PR	27-SEP-2000; 2000US-0236359.	
PR	04-OCT-2000; 2000GB-0024263.	
XX		
PA	(MOLE-)	MOLECULAR DYNAMICS INC.

xx	Penn SG, Hanzel DK, Chen W, Rank DR;
xx	WPI; 2001-488899/53.
DR	
xx	
pt	Single exon nucleic acid probes for analyzing gene expression in human
pr	hearts -
xx	
ps	Claim 4; SEQ ID No 18226; 530bp; English.
xx	
cc	The present invention relates to single exon nucleic acid probes for
cc	measuring human gene expression in a sample derived from human heart. The
cc	present sequence is one such probe. The probes may be used for
cc	predicting, measuring and displaying gene expression in samples derived
cc	from the human heart via microarrays. By measuring gene expression, the
cc	probes are useful for predicting, diagnosing, grading, staging,
cc	monitoring and prognosing diseases of the human heart and vascular system
cc	e.g. cardiovascular disease, hypertension, cardiac arrhythmias and
cc	congenital heart disease.
cc	Note: The sequence data for this patent did not form part of the printed
cc	specification, but was obtained in electronic format directly from WIPO
cc	at ftp.wipo.int/pub/published_pct_sequences.
xx	
sq	Sequence 273 BP; 80 A; 36 C; 81 G; 76 T; 0 other;
	Query Match 14.7%; Score 58.4; DB 22; Length 273:
	Best Local Similarity 74.0%; Pred. No. 7.6e-09;
	Matches 74; Conservative 0; Mismatches 26; Indels 0; Gaps 0
Oy	297 GGCGAATATTGGCTAAATGAGAGAACAATTTTGGCTAATAAGGAAGAATTTCTTAAGCA 356
Dd	115 GGGGAACATTCAGCATTCAGAGAGATCATTGGGTCTGTTAAAGCA 174
Oy	357 GCAAAGCATTCAGAGAGTGACTTGGGTGCTGTTAAAGCA 396
Dd	175 GCAAAGCATTCAGAGAGTGACTTGGGTGCTGTTAAAGCA 214
RESULT 5	
AKK23591	
ID	AKK23591 standard; DNA; 273 BP.
XX	
AC	AKK23591;
XX	
DT	05-NOV-2001 (first entry)
XX	
DE	Human brain expressed single exon probe SEQ ID NO: 23582.
XX	
KM	Human; brain expressed exon; gene expression analysis; probe;
KW	microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;
XX	epilepsy; cancer; ss.
OS	Homo sapiens.
XX	
PN	WO200157275-A2.
XX	
PD	09-AUG-2001.
XX	
PE	30-JAN-2001; 2001WO-US00667.
XX	
PR	04-FEB-2000; 2000US-0180312.
PR	26-MAY-2000; 2000US-0207456.
PR	30-JUN-2000; 2000US-0608408.
PR	03-AUG-2000; 2000US-0632365.
PR	21-SEP-2000; 2000US-0234687.
PR	27-SEP-2000; 2000US-0236359.
PR	04-OCT-2000; 2000GB-0024263.
XX	
PA	(MOLE-) MOLECULAR DYNAMICS INC.
XX	
PT	Penn SG, Hanzel DK, Chen W, Rank DR;
XX	WPI; 2001-483446/52.
DR	

[illegible]

CC	such as lymphoma, leukaemia and myeloma. The present sequence is one of
CC	the probes of the invention.
XX	
XX	Sequence 273 BP; 80 A; 36 C; 81 G; 76 T; 0 other:
XX	
XX	Query Match 14.7%; Score 58.4; DB 22; Length 273;
XX	Best Local Similarity 74.0%; Pred. No. 7.6e-09;
XX	Matches 74; Conservative 0; Mismatches 26; Indels 0; Gaps
OY	297 GGTGAATATATTTGGCTAAATGAGGAACAATTTTGGCTAATAAGAGAAATTTCTAAGCA 356
DB	115 GGGGAACATTGAACTTGAGAGAGATGATTTGGGCTATCTGTTGGAATTTCTAAACA 174
OY	357 GCAAGCATTCAAGAGCTGAGCTGGCGCTTAAAGCA 396
DB	175 GCAAGCATTCAAGAGCTGAGCTGGCTGCTTAAAGCA 214
RESULT 7	
ID	AA126826 standard; DNA: 273 BP.
XX	AA126826;
XX	12-Oct-2001 (first entry)
DE	Probe #16759 for gene expression analysis in human cervical cell sample.
XX	
KW	Probe; human; microarray; gene expression; cervical epithelial cell;
KW	cervical cancer; ss.
XX	
OS	Homo sapiens.
XX	
PN	WO200157278-A2.
PD	09-AUG-2001.
PF	30-JAN-2001; 2001WO-US00670.
XX	
PR	04-FEB-2000; 2000US-0180312.
PR	26-MAY-2000; 2000US-0207456.
PR	30-JUN-2000; 2000US-0608408.
PR	03-AUG-2000; 2000US-0632366.
PR	21-SEP-2000; 2000US-0234687.
PR	27-SEP-2000; 2000US-0236359.
PR	04-OCT-2000; 2000GB-0024263.
PA	(MOLE-) MOLECULAR DYNAMICS INC.
XX	
PI	Penn SG, Hanzel DK, Chen W, Rank DR;
DR	WPI: 2001-488901/53.
XX	
PT	Human genome-derived single exon nucleic acid probes useful for
PT	analyzing gene expression in human cervical epithelial cells -
XX	
PS	Claim 25; SEQ ID NO 16759; 487bp; English.
CC	
CC	The present invention relates to human single exon nucleic acid probes
CC	(SENPs). The present sequence is one such probe. The SENPs are derived
CC	from human HeLa cells. The SENPs can be used to produce a single exon
CC	microarray, which can be used for measuring human gene expression in a
CC	sample derived from human cervical epithelial cells. By measuring gene
CC	expression, the probes are therefore useful in grading and/or staging
CC	of diseases of the cervix, notably cervical cancer.
CC	Note: The sequence data for this patent did not form part of the printed
CC	specification, but was obtained in electronic format directly from WIPO
CC	at ftp.wipo.int/pub/published_pct_sequences .
XX	
XX	Sequence 273 BP; 80 A; 36 C; 81 G; 76 T; 0 other:
XX	
XX	Query Match 14.7%; Score 58.4; DB 22; Length 273;
XX	Best Local Similarity 74.0%; Pred No. 7.6e-09;

RESULT 9

XX

FIGURE 6 | Gene expression analysis in human heart cells sample.

FIGURE 6 | Gene expression analysis in human heart cells sample.

KW Human; gene expression; heart; microarray; vascular system; probe;
 KW cardiovascular disease; hypertension; cardiac arrhythmia;
 KW congenital heart disease; ss.
 OS Homo sapiens.
 XX WO200157274-A2.
 PN 09-AUG-2001.
 XX 30-JAN-2001; 2001WO-US00666.
 XX 04-FEB-2000; 2000US-0180312.
 XX 26-MAY-2000; 2000US-0207456.
 XX 30-JUN-2000; 2000US-0608408.
 XX 03-AUG-2000; 2000US-0632366.
 XX 21-SEP-2000; 2000US-0234687.
 XX 27-SEP-2000; 2000US-0236359.
 XX 04-OCT-2000; 2000GB-0024263.
 PA (MOLE-) MOLECULAR DYNAMICS INC.
 XX Penn SG, Hanzel DK, Chen W, Rank DR;
 PI WPI; 2001-488899/53.
 XX Single exon nucleic acid probes for analyzing gene expression in human hearts -
 PT
 PT
 XX Claim 1; SEQ ID No 8349; 530pp; English.
 XX The present invention relates to single exon nucleic acid probes for measuring human gene expression in a sample derived from human heart. The present sequence is one such probe. The probes may be used for predicting, measuring and displaying gene expression in samples derived from the human heart via microarrays. By measuring gene expression, the probes are useful for predicting, diagnosing, grading, staging, monitoring and prognosing diseases of the human heart and vascular system e.g. cardiovascular disease, hypertension, cardiac arrhythmias and congenital heart disease.
 CC Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO CC at ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 519 BP; 170 A; 68 C; 148 G; 133 T; 0 other;
 Query Match 14.7%; Score 58.4; DB 22; Length 519;
 Best Local Smilarity 74.0%; Pred. No. 1.1e-08;
 Matches 74; Conservative 0; Mismatches 26; Indels 0; Gaps 0;
 QY 297 GGTGAATATTGGCTAATGAGGAGCAAAATTTGGCTAATAAGGAGAAATTTCTTAAGCA 356
 DB 254 GGGGACATTGAACCTGAGAGAGATGATTTGGGGTATCTGCTGGGAGAAATTTCTTAACA 313
 QY 357 GCAAAGCATTCAAGAGTGACTTGGGTGCTGCTTTAAAGGCA 396
 DB 314 GCAAAGCATTCAAGAGTGACTTGGGTGCTGCTTTAAAGGCA 353
 RESULT 11
 AAK10911
 ID AAK10911 standard; DNA; 519 BP.
 XX
 AC AAK10911;
 XX
 DT 05-NOV-2001 (first entry)
 XX
 DE Human brain expressed single exon probe SEQ ID NO: 10902.
 XX
 KW Human; brain expressed exon; gene expression analysis; probe;
 KW microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;
 KW epilepsy; cancer; ss.
 XX

OS Homo sapiens.
 XX WO200157275-A2.
 XX 09-AUG-2001.
 XX 30-JAN-2001; 2001WO-US00667.
 XX 04-FEB-2000; 2000US-0180312.
 XX 26-MAY-2000; 2000US-0207456.
 XX 30-JUN-2000; 2000US-0608408.
 XX 03-AUG-2000; 2000US-0632366.
 XX 21-SEP-2000; 2000US-0234687.
 XX 27-SEP-2000; 2000US-0236359.
 XX 04-OCT-2000; 2000GB-0024263.
 PA (MOLE-) MOLECULAR DYNAMICS INC.
 XX Penn SG, Hanzel DK, Chen W, Rank DR;
 PI WPI; 2001-483446/52.
 XX Single exon nucleic acid probes for analyzing gene expression in human brains -
 PT
 PT
 XX Example 4; SEQ ID NO: 10902; 650pp + Sequence Listing; English.
 XX The present invention provides a number of single exon nucleic acid probes which are derived from genomic sequences expressed in the human brain. They can be used to measure gene expression in brain cell samples, which may enable the diagnosis and improved treatment of nervous system diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia, epilepsy and cancers. The present sequence is one of the probes of the invention.
 CC
 SQ Sequence 519 BP; 170 A; 68 C; 148 G; 133 T; 0 other;
 Query Match 14.7%; Score 58.4; DB 22; Length 519;
 Best Local Similarity 74.0%; Pred. No. 1.1e-08;
 Matches 74; Conservative 0; Mismatches 26; Indels 0; Gaps 0;
 QY 297 GGTGAATATTGGCTAATGAGGAGCAAAATTTGGCTAATAAGGAGAAATTTCTTAAGCA 356
 DB 254 GGGGACATTGAACCTGAGAGAGATGATTTGGGGTATCTGCTGGGAGAAATTTCTTAACA 313
 QY 357 GCAAAGCATTCAAGAGTGACTTGGGTGCTGCTTTAAAGGCA 396
 DB 314 GCAAAGCATTCAAGAGTGACTTGGGTGCTGCTTTAAAGGCA 353
 RESULT 12
 AAK36769
 ID AAK36769 standard; DNA; 519 BP.
 XX
 AC AAK36769;
 XX
 DT 06-NOV-2001 (first entry)
 XX
 DE Human bone marrow expressed single exon probe SEQ ID NO: 11326.
 XX
 KW Human; bone marrow expressed exon; gene expression analysis; probe;
 KW microarray; cancer; leukaemia; lymphoma; myeloma; ss.
 XX
 OS Homo sapiens.
 XX WO200157276-A2.
 XX 09-AUG-2001.
 XX 30-JAN-2001; 2001WO-US00668.
 XX 04-FEB-2000; 2000US-0180312.
 XX 26-MAY-2000; 2000US-0207456.
 PR

```
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-488900/53.
XX Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human bone marrow -
XX
XX Example 4; SEQ ID NO: 11326; 658pp + Sequence Listing; English.
XX
XX The present invention provides a number of single exon nucleic acid
CC probes which are derived from genomic sequences expressed in the human
CC bone marrow. They can be used to measure gene expression in bone marrow
CC samples, which may enable the improved diagnosis and treatment of cancers
CC such as lymphoma, leukemia and myeloma. The present sequence is one of
CC the probes of the invention.
XX
XX Sequence 519 BP; 170 A; 68 C; 148 G; 133 T; 0 other;
SQ
Query Match 14.7%; Score 58.4; DB 22; Length 519;
Best Local Similarity 74.0%; Pred. No. 1.1e-08;
Matches 74; Conservative 0; Mismatches 26; Indels 0; Gaps 0;
QY 297 GGTGAATATTTGGCTAATGAGGAGCAAAATTTGGCTAATAGGAAGAAATTTCTAAGCA 356
DB 254 GGGGAACATTGAACTTGAGAGAGATGATTTGGGGTATCTGTTGAAGAAATTTCTAAGCA 313
QY 357 GCAAAGCATTCAAGAGTGCTGCTGGTCTGTTAAAGGCA 396
DB 314 GCAAAGCATTCAAGAGTGCTGCTGGTCTGTTAAAGGCA 353
RESULT 13
AAI17621
ID AAI17621 standard; DNA; 519 BP.
XX
XX AAI17621;
XX
DT 12-OCT-2001 (first entry)
XX
DE Probe #7554 for gene expression analysis in human cervical cell sample.
XX Probe; human; microarray; gene expression; cervical epithelial cell;
KW cervical cancer; ss.
XX Homo sapiens.
XX
XX WO2001572728-A2.
XX
XX 09-AUG-2001.
XX
XX 30-JAN-2001; 2001WO-US000670.
XX
XX 04-FEB-2000; 2000US-0180312.
XX 26-MAY-2000; 2000US-0207456.
XX 30-JUN-2000; 2000US-0608408.
XX 03-AUG-2000; 2000US-0632366.
XX 21-SEP-2000; 2000US-0234687.
XX 27-SEP-2000; 2000US-0236359.
XX 04-OCT-2000; 2000GB-0024263.
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-488901/53.
XX
```

```
XX Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human cervical epithelial cells -
XX
XX Cialm 25; SEQ ID No 7554; 487pp; English.
XX
XX The present invention relates to human single exon nucleic acid probes
CC (SENP). The present sequence is one such probe. The SENPs are derived
CC from human HeLa cells. The SENPs can be used to produce a single exon
CC microarray, which can be used for measuring human gene expression in a
CC sample derived from human cervical epithelial cells. By measuring gene
CC expression, the probes are therefore useful in grading and/or staging
CC of diseases of the cervix, notably cervical cancer.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 519 BP; 170 A; 68 C; 148 G; 133 T; 0 other;
SQ
Query Match 14.7%; Score 58.4; DB 22; Length 519;
Best Local Similarity 74.0%; Pred. No. 1.1e-08;
Matches 74; Conservative 0; Mismatches 26; Indels 0; Gaps 0;
QY 297 GGTGAATATTTGGCTAATGAGGAGCAAAATTTGGCTAATAGGAAGAAATTTCTAAGCA 356
DB 254 GGGGAACATTGAACTTGAGAGAGATGATTTGGGGTATCTGTTGAAGAAATTTCTAAGCA 313
QY 357 GCAAAGCATTCAAGAGTGCTGCTGGTCTGTTAAAGGCA 396
DB 314 GCAAAGCATTCAAGAGTGCTGCTGGTCTGTTAAAGGCA 353
RESULT 14
AAI42541
ID AAI42541 standard; DNA; 519 BP.
XX
XX AAI42541;
XX
DT 17-OCT-2001 (first entry)
XX
DE Probe #11227 used to measure gene expression in human placenta sample.
XX Probe; microarray; human; placenta; antenatal diagnosis;
KW genetic disorder; ss.
XX Homo sapiens.
XX
XX WO200157272-A2.
XX
XX 09-AUG-2001.
XX
XX 30-JAN-2001; 2001WO-US000663.
XX
XX 04-FEB-2000; 2000US-0180312.
XX 26-MAY-2000; 2000US-0207456.
XX 30-JUN-2000; 2000US-0608408.
XX 03-AUG-2000; 2000US-0632366.
XX 21-SEP-2000; 2000US-0234687.
XX 27-SEP-2000; 2000US-0236359.
XX 04-OCT-2000; 2000GB-0024263.
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-488897/53.
XX Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human placenta -
XX
XX Cialm 25; SEQ ID No 11227; 654pp; English.
XX
XX The present invention relates to single exon nucleic acid probes (SENP).
```


GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 19, 2003, 23:16:26 ; Search time 1262.56 Seconds
(without alignments)
9128.055 Million cell updates/sec

Title: US-09-997-610-1_COPY_53_448

Perfect score: 396

Sequence: 1 gctggacctccagcaccc.....cttgggtgctgttaaaaggca 396

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.*

1: gb_ba.*

2: gb_htg.*

3: gb_in.*

4: gb_om.*

5: gb_ov.*

6: gb_pat.*

7: gb_ph.*

8: gb_pl.*

9: gb_pr.*

10: gb_ro.*

11: gb_sts.*

12: gb_sy.*

13: gb_un.*

14: gb_vi.*

15: em_ba.*

16: em_fun.*

17: em_hum.*

18: em_in.*

19: em_mu.*

20: em_om.*

21: em_or.*

22: em_ov.*

23: em_pat.*

24: em_ph.*

25: em_pl.*

26: em_ro.*

27: em_sts.*

28: em_un.*

29: em_vi.*

30: em_htg_hum.*

31: em_htg_inv.*

32: em_htg_other.*

33: em_htg_mus.*

34: em_htg_pln.*

35: em_htg_rod.*

36: em_htg_mam.*

37: em_htg_vrt.*

38: em_sy.*

39: em_htgo_hum.*

40: em_htgo_mus.*

41: em_htgo_other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	351	88.6	145880	9	HS302D9	282198 Human DNA s
2	146.2	36.9	1265	10	TMSHP20A	DI2974 Tamias asia
3	139.2	35.2	5121	10	AB067779	AB067779 Tamias si
4	83	21.0	2005	10	TMSHP25	DI2975 Tamias asia
5	68.6	17.3	559	10	AB067813S3	AB069679 Callosclu
6	68.2	17.2	145880	9	HS302D9	282198 Human DNA s
7	63.4	16.0	90497	9	HS73284	AL008722 Human DNA
8	63.4	16.0	135880	2	AC104256	AC104256 Homo sapi
9	63.2	16.0	51000	9	AP005265	AP005265 Homo sapi
10	63.2	16.0	144514	2	AC027433	AC027433 Homo sapi
11	63.2	16.0	149038	2	AC025112	AC025112 Homo sapi
12	62.6	15.8	162063	9	AC015553	AC015553 Homo sapi
13	62.6	15.8	175134	2	AL161623	AL161623 Homo sapi
14	62.6	15.8	176932	9	AL157884	AL157884 Human DNA
15	62.4	15.8	1385	10	TMSHP27	DI2976 Tamias asia
16	62	15.7	203234	9	AL441943	AL441943 Human DNA
17	61.8	15.6	69937	2	AC069566	AC069566 Homo sapi
18	61.8	15.6	122231	2	AC026099	AC026099 Homo sapi
19	61.8	15.6	143872	9	AC091987	AC091987 Homo sapi
20	61.8	15.6	172792	9	AC062031	AC062031 Homo sapi
21	61.8	15.6	179848	9	AC092967	AC092967 Homo sapi
22	61.8	15.6	180038	2	AC026870	AC026870 Homo sapi
23	61.2	15.5	177386	2	AC018971	AC018971 Homo sapi
24	60.2	15.2	108511	9	AC011418	AC011418 Homo sapi
25	60.2	15.2	137591	2	AL672086	AL672086 Homo sapi
26	60.2	15.2	158587	9	AC074325	AC074325 Homo sapi
27	60.2	15.2	158811	2	AC013503	AC013503 Homo sapi
28	60.2	15.2	160151	9	AC098595	AC098595 Homo sapi
29	60.2	15.2	161049	2	AL591842	AL591842 Homo sapi
30	60.2	15.2	173767	9	AC002449	AC002449 Human PAC
31	60.2	15.2	175765	9	AC027128	AC027128 Homo sapi
32	60.2	15.2	179484	9	AC107626	AC107626 Homo sapi
33	60.2	15.2	187201	2	AC068063	AC068063 Homo sapi
34	60.2	15.2	187640	9	AC073840	AC073840 Homo sapi
35	60.2	15.2	188622	9	AL359258	AL359258 Human DNA
36	60.2	15.2	192973	9	AC023471	AC023471 Homo sapi
37	60.2	15.2	207600	9	AC024910	AC024910 Homo sapi
38	60.2	15.2	222542	9	AC023379	AC023379 Homo sapi
39	59.8	15.1	170200	9	AC008012	AC008012 Homo sapi
40	59.6	15.1	144057	2	AC074266	AC074266 Homo sapi
41	59.6	15.1	172349	9	AC019185	AC019185 Homo sapi
42	59.6	15.1	195594	2	AC096992	AC096992 Homo sapi
43	59.6	15.1	197078	2	AC117382	AC117382 Homo sapi
44	59.6	15.1	233877	9	AC093798	AC093798 Homo sapi
45	59.4	15.0	98070	9	AC025463	AC025463 Homo sapi

ALIGNMENTS

RESULT 1
HS302D9
LOCUS
DEFINITION Human DNA sequence from clone RP1-302D9 on chromosome 22 Contains
GSSs, complete sequence.
ACCESSION 282198
VERSION 282198.2
KEYWORDS HTG.
SOURCE
ORGANISM Homo sapiens.
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 145880)
AUTHORS Bridgeman,A.
TITLE Direct Submission

HS302D9 145880 bp DNA linear PRI 12-DEC-1999
Human DNA sequence from clone RP1-302D9 on chromosome 22 Contains
GSSs, complete sequence.
282198 GI:6572207
HTG.
Homo sapiens
Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 145880)
Bridgeman,A.
Direct Submission

JOURNAL

Submitted (08-DEC-1999) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk

COMMENT

On Dec 13, 1999 this sequence version replaced gi:3164067. During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

This sequence has been finished according to sequence map criteria as follows. An attempt is made to resolve all sequencing problems, such as compressions and repeats, but not necessarily within known annotated human repeat sequence elements (e.g. Alu). Where the sequence is ambiguous, there is an annotation using the 'unsure' feature key.

The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP database can be found at

<http://www.sanger.ac.uk/projects/C.elegans/wormpep> This sequence was generated from part of bacterial clone contigs of human chromosome 22, constructed by the Sanger Centre Chromosome 22 Mapping Group. Further information can be found at

<http://www.sanger.ac.uk/HGP/Chr22>
RPI-302D9 is from the library RPCI-1 constructed at the Roswell Park Cancer Institute by the group of Pieter de Jong. For further details see <http://bacpac.med.buffalo.edu/> VECTOR: PCYPAC2

This sequence is the entire insert of clone RPI-302D9 The true left end of clone CTA-282F2 is at 69682 in this sequence. The true right end of clone CTA-41562 is at 55167 in this sequence.

FEATURES

Source

1..145880
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="22"
/clone="RPI-302D9"
/clone_lib="RPCI-1"

repeat_region

/note="MER3 repeat: matches 144. .209 of consensus"
246. .571

repeat_region

/note="AluSx repeat: matches 1. .312 of consensus"

repeat_region

/note="MER3 repeat: matches 1. .144 of consensus"

repeat_region

/note="MER5A repeat: matches 26. .187 of consensus"
1033. .1336

repeat_region

/note="AluSp repeat: matches 1. .299 of consensus"

repeat_region

1450. .1583
/note="MIR repeat: matches 24. .160 of consensus"

repeat_region

1687. .1752
/note="L2 repeat: matches 2593. .2661 of consensus"
2350. .2660

repeat_region

/note="AluSc repeat: matches 3. .309 of consensus"

repeat_region

2684. .2981
/note="AluSq repeat: matches 2. .300 of consensus"

repeat_region

3323. .3343
/note="MLTIE repeat: matches 116. .136 of consensus"

repeat_region

3344. .3652
/note="AluY repeat: matches 1. .309 of consensus"

repeat_region

3653. .3928
/note="MLTIE repeat: matches 136. .359 of consensus"

repeat_region

3929. .4278
/note="THE1B repeat: matches 3. .364 of consensus"

repeat_region

4279. .4485
/note="MLTIE repeat: matches 359. .568 of consensus"

repeat_region

5073. .5176
/note="52 copies 2 mer ct 78 conserved"

repeat_region

5181. .5491
/note="AluJb repeat: matches 1. .311 of consensus"

repeat_region

6369. .6485
/note="L2 repeat: matches 2579. .2705 of consensus"

repeat_region

repeat_region

6647. .6685
/note="MADE1 repeat: matches 1. .23 of consensus"

repeat_region

6686. .6987
/note="AluSx repeat: matches 1. .302 of consensus"

repeat_region

6988. .7036
/note="MADE1 repeat: matches 23. .77 of consensus"

repeat_region

7482. .7754
/note="AluJb repeat: matches 9. .290 of consensus"

repeat_region

7775. .8060
/note="AluJo repeat: matches 1. .295 of consensus"

repeat_region

8414. .8551
/note="L2 repeat: matches 2553. .2706 of consensus"

repeat_region

8914. .9030
/note="MIR repeat: matches 147. .262 of consensus"

repeat_region

9110. .9280
/note="MIR repeat: matches 91. .262 of consensus"

repeat_region

9283. .9412
/note="MIR repeat: matches 15. .144 of consensus"

repeat_region

9521. .9679
/note="FAM repeat: matches 3. .161 of consensus"

repeat_region

9820. .10225
/note="MSTB repeat: matches 2. .425 of consensus"

misc_feature

complement(10179. .10678)

misc_feature

/note="match: GSS: Em:B56592"

misc_feature

/note="match: GSS: Em:AQ701486"

misc_feature

complement(10249. .10706)

repeat_region

/note="match: GSS: Em:AQ225495"

misc_feature

10312. .10383
/note="MIR repeat: matches 79. .150 of consensus"

misc_feature

10718. .11310
/note="match: GSS: Em:B14024"

misc_feature

10784. .11201
/note="match: GSS: Em:B43656"

repeat_region

11838. .11946
/note="MIR repeat: matches 20. .137 of consensus"

repeat_region

12174. .12445
/note="L2 repeat: matches 1988. .2275 of consensus"

repeat_region

12444. .12642
/note="MIR repeat: matches 63. .241 of consensus"

misc_feature

13017. .13369
/note="match: STS: Em:G49301"

repeat_region

13331. .13397
/note="MIR repeat: matches 174. .244 of consensus"

repeat_region

13398. .13698
/note="AluSp repeat: matches 1. .302 of consensus"

repeat_region

13699. .13810
/note="MIR repeat: matches 76. .174 of consensus"

repeat_region

13806. .13919
/note="MIR repeat: matches 77. .189 of consensus"

repeat_region

13945. .14060
/note="MIR repeat: matches 24. .142 of consensus"

repeat_region

14061. .14367
/note="AluY repeat: matches 1. .301 of consensus"

repeat_region

14368. .14452
/note="MIR repeat: matches 141. .225 of consensus"

repeat_region

14589. .14679
/note="MIR repeat: matches 173. .262 of consensus"

misc_feature

14597. .15201
/note="match: GSS: Em:AQ553482"

misc_feature

14616. .15060
/note="match: GSS: Em:AQ370601"

repeat_region

14868. .15040
/note="MIR repeat: matches 49. .233 of consensus"

repeat_region

15071. .15188
/note="L2 repeat: matches 2112. .2239 of consensus"

repeat_region

15304. .15399
/note="MLT1B repeat: matches 1. .99 of consensus"

repeat_region

15490. .15662
/note="AluS91 repeat: matches 2. .114 of consensus"

repeat_region

15669. .15727
/note="MLT1B repeat: matches 119. .178 of consensus"

repeat_region

15728. .16027

repeat_region /note="AluSc repeat: matches 1. .299 of consensus" 16028. .16245
repeat_region /note="MLT1B repeat: matches 178. .390 of consensus" 16546. .16854
repeat_region /note="AluY repeat: matches 1. .300 of consensus" 18296. .18323
repeat_region /note="MSTA repeat: matches 2. .29 of consensus" 18324. .18392
repeat_region /note="MER66-internal repeat: matches 4919. .4993 of consensus" 18393. .18712
repeat_region /note="AluJb repeat: matches 1. .311 of consensus" 18713. .19133
repeat_region /note="MER66-internal repeat: matches 4548. .4919 of consensus" 19251. .19719
misc_feature complement(18872. .19230)
misc_feature /note="match: GSS: Em:AQ005063"
repeat_region /note="match: GSS: Em:B14179"
repeat_region /note="HERVH21 repeat: matches 4657. .5784 of consensus" 20317. .20382
repeat_region /note="33 copies 2 mer ta 68 conserved" 20513. .20666
repeat_region /note="77 copies 2 mer tt 70 conserved" 20682. .21008
repeat_region /note="AluSg1 repeat: matches 1. .306 of consensus" 21239. .21553
repeat_region /note="HUEB-P3 repeat: matches 4410. .4713 of consensus" 21882. .22254
repeat_region /note="THEB1B repeat: matches 1. .364 of consensus" 22302. .22537
repeat_region /note="MER66-internal repeat: matches 2186. .2417 of consensus" 22538. .22850
repeat_region /note="AluSg repeat: matches 1. .313 of consensus" 22851. .23801
repeat_region /note="MER66-internal repeat: matches 1210. .2186 of consensus" 23905. .23989
repeat_region /note="MER66-internal repeat: matches 3017. .3102 of consensus" 23905. .23989

Query Match 88.6%; Score 351; DB 9; Length 145880;
Best Local Similarity 98.6%; Pred. No. 2.4e-94;
Matches 354; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 38 TGGGGCCTCTGTGCACCAAGTTTACCACAATATACAGGAGAAATAAGTGAAATGACAA 97
Db 37254 TGAATGTCTGTGTCCTGATATAGAAAGTTCAGCCTTTACTGTGAAGCTCAGTGGAAACTTTC 157
QY 98 AATGCCCTCTCTGTATAGAAAGTTCAGCCTTTACTGTGAAGCTCAGTGGAAACTTTC 157
Db 37314 AATGCCCTCTCTGTATAGAAAGTTCAGCCTTTACTGTGAAGCTCAGTGGAAACTTTC 37373
QY 158 CTCCTCTTCAAGCCCATCATCTTCACAGGGTCTCTGTACAAATGCCAGAGGGATTAA 217
Db 37374 CTCCTCTTCAAGCCCATCATCTTCACAGGGTCTCTGTACAAATGCCAGAGGGATTAA 37433
QY 218 AGAGGCCATGGAGTCTTGTCTGTCAGGTCCTGGGAATTACTACTCCAGCTTTGATG 277
Db 37434 AGAGGCCATGGAGTCTTGTCTGTCAGGTCCTGGGAATTACTACTCCAGCTTTGATG 37493
QY 278 TTGAGCTGCATCTTCAAGAGTGAATATTTGGCTAATGAGGAAGCAAAATTTTGGCTAATA 337
Db 37494 TTGAGCTGCATCTTCAAGAGTGAATATTTGGCTAATGAGGAAGCAAAATTTTGGCTAATA 37553
QY 338 AGAAGAAATTTCTAAGACGAAAGCAATTTCAAGAGGTGACTTGGGTGCTGTTAAAGGCA 396
Db 37554 AGAAGAAATTTCTAAGACGAAAGCAATTTCAAGAGGTGACTTGGGTGCTGTTAAAGGCA 37612

RESULT 2
TMSHP20A

LOCUS TMSHP20A 1265 bp mRNA linear ROD 03-FEB-1999
DEFINITION Tamias asiaticus mRNA for HP-20, complete cds.
ACCESSION D12974
VERSION D12974.1 GI:287467
KEYWORDS HP-20; collagen-like domain; hibernation-related protein; plasma protein.
SOURCE Tamias asiaticus liver cDNA to mRNA, clone:pCM20-7.
ORGANISM Tamias sibiricus
Mammalia; Eutheria; Rodentia; Sciurognathi; Scuridae; Scurinae; Tamias.
REFERENCE 1 (bases 1 to 1265)
AUTHORS Takamatsu,N., Ohba,K., Kondo,J., Kondo,N. and Shiba,T.
TITLE Hibernation-associated gene regulation of plasma proteins with a collagen-like domain in mammalian hibernators
JOURNAL Mol. Cell. Biol. 13 (3), 1516-1521 (1993)
REFERENCE 2 (bases 1 to 1265)
AUTHORS Takamatsu,N.
TITLE Direct Submission
JOURNAL Submitted (21-AUG-1992) Nobuhiko Takamatsu, School of Hygienic Sciences, Kitasato University, Laboratory of Molecular Biology; 1-15-1 Kitasato, Sagamihara, Kanagawa 228, Japan (Tel:0427-78-9408, Fax:0427-78-9403)
FEATURES
Location/Qualifiers
1..1265
/organism="Tamias sibiricus"
/db_xref="taxon:64680"
/clone="pCM20-7"
/tissue_type="liver"
89..679
/note="collagen-like domain at nt 161-277"
/codon_start=1
/product="HP-20"
/protein_id="BAA02351.1"
/db_xref="GI:287468"
/translation="MTDVMRLAIFVLNVLNDVSCSGPPGVGVPVGPVGPGRGP
PQPGAAGRPDGPVKPSVKPCRSAPTKVFSRPLPPVSEPVVFTLVNTQDRL
KESTGVNCEPVGNYHFSFDVELYHCKVKIQLMKNHIOVMKHLQSKNEYENASGAMI
MPLRGDKVWLEADVETEPDQAKVIVFSGFLISS"
89..157
sig_peptide
158..676
mat_peptide
/product="HP-20"
polyA_signal
1238..1243
polyA_site
1265
BASE COUNT 354 a 304 c 288 g 319 t
ORIGIN
Query Match 36.9%; Score 146.2; DB 10; Length 1265;
Best Local Similarity 68.0%; Pred. No. 7e-33;
Matches 221; Conservative 0; Mismatches 98; Indels 6; Gaps 1;
QY 2 CTGGACCTCCAGCACACCCAGGCCCCAGAGAAAGTGGGGCCTCTCTGGTGACACAGGTT 61
Db 186 CAGGAGTGGCCGAGTGGCCGGGCTCTGTGGCCCCCAGGTCAACGAGCGCTGCGGGTA 245
QY 62 TACCAACAATATACAGAGAAATAAGTGAAATGACAAATGCCCTCTGTCTGATATAGAAA 121
Db 246 GGCCAGGAGACCAGGACCAAAAGGCCAGTGTAAATGCCCTGCG-----AGAGAGA 299
QY 122 GGTACGCCCTTTACTGTGAAGCTCAGTGGAAACTTCTCTTCTTTCAAGCCCATCATCT 181
Db 300 GGTACGCCCTTACGGTGAAGTTCAGCGGAAGGCTCCCTCCACCTTCGGAGCCTGTGTCT 359
QY 182 TCACAGGGTCTGTACAAATGCCAGAGGATTTAAGGAGGCATGGGAGTCTTTGCTT 241
Db 360 TCACAGAGTCTGTACAAATGCCAGGAGACTTGAAGGAGACACAGGAGTCTTTAACT 419
QY 242 GCAGGGCTGGGAATTTACTACTCCAGCTTTTGCATGTTGAGTGCATCATTTGCAAGTGA 301
Db 420 GCGTGGAGCTGGAAATTTACCATTTTCAGCTTTGATGTTGAGCTTTACCAGTGCAGGTGA 479
QY 302 ATATTTGGCTAATGAGGAAGCAAAAT 326

```

Db 480 AGATTGGTTGATGAAGAACACAT 504
|||||
RESULT 3
AB067779 5121 bp DNA linear ROD 26-OCT-2001
LOCUS Tamias sibiricus gene for HP-20, complete cds.
DEFINITION AB067779
ACCESSION AB067779.1 GI:15706341
VERSION
KEYWORDS
SOURCE Tamias sibiricus DNA, clone:lambdaCM20G1.
ORGANISM Tamias sibiricus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Scuridae; Sciurinae;
Tamias.
REFERENCE
1 Ono,M., Hosoe,Y., Azuma,S., Shoji,M., Nara,K., Kondo,N., Shiba,T.
and Takamatsu,N.
HNF-1 regulates the liver-specific transcription of the chipmunk
HP-20 gene
Gene 277 (1-2), 121-127 (2001)
21488336
2 (bases 1 to 5121)
Ono,M.
Direct Submission
Submitted (04-AUG-2001) Motoharu Ono, Kitasato University,
Department of Biosciences, School of Science; 1-15-1 Kitasato,
Sagamihara, Kanagawa 228-8555, Japan
(E-mail:ms9805@estu.sci.kitasato-u.ac.jp, Tel:81427789408,
Fax:81427789408)
FEATURES
source
1..5121
/organism="tamias sibiricus"
/db_xref="taxon:64680"
/clone="lamdaCM20G1"
/notes="synonym:Tamias asiaticus"
1518..1684
join(1678..1684,2573..2710,3968..4413)
/gene="HP-20"
join(1678..1684,2573..2710,3968..4413)
/gene="HP-20"
/codon_start=1
/product="HP-20"
/protein_id="BAB68362.1"
/db_xref="GI:15706342"
/translation="MTDAWRLAIFVLVWVNDQVSCSPGPVGYPGVPGPRGP
PQOPGAAGRPDGPGRKPSVKPCPCRSFTKFSGRLPPEPVVFTVLYNTQDL
KASTGYNCFVEPNHFSFDVELYHCVKVIGLMKNHIQYMEKHQLSKNEYENASGAMI
MPLRQGDVWLEADVETEPDQAKVYVFSGLISS"
2573..2710
exon
/gene="HP-20"
3968..5000
exon
/gene="HP-20"
BASE COUNT 1571 a 1018 c 1090 g 1442 t
ORIGIN
Query Match 35.2%; Score 139.2; DB 10; Length 5121;
Best Local Similarity 71.6%; Pred. No. 9.9e-31;
Matches 199; Conservative 0; Mismatches 73; Indels 6; Gaps 1;
QY 49 GGTGCACCAAGTTTACCACAATATACAGGAGAATAAGTGAAATGACAAAATGCCCTGT 108
Db 3967 GCGCTCGGTAGGCAGGAGCCAGACCAAGGCCCAAGTGTAAATGCCCTG- 4025
QY 109 CTTGATATAGAAAGTCAAGCTTTTACTGTGAAGCTCAGTGGAAACTTCCTCTTCCTTC 168
Db 4026 -----CAGAGAGAGGTCAAGCTTACGCGTGAAGTTTACGCGGAAGGCTCCCTCCACCTTCG 4080
QY 169 AAGCCCATCATCTTACAGAGGGTCCCTGTACAATGCCAGAGGATTTAAAGAGGGCATG 228
Db 4081 GAGCCTGTGCTTTCACAGAGGTCCTGTACATACCCAGAGGGACTTGAAGGCGGACACA 4140
QY 229 GGAGTCTTTGCTTGCAGGGTGCCTGGGAATTACTACTCCAGCTTTGATGTGAGTGCAT 288
|||||
Db 4141 GGAGTCTTTAACTGCGTGGAGCCTGGAATATACCATTTAGCTTGTGATGCTTTAC 4200
|||||
QY 289 CATTGCAAGGTGAATATTTGGCTAATGAGGAACAAT 326
|||||
Db 4201 CACTGCAAGTGAAGATTGGGTTGATGAAGAACACAT 4238
|||||
RESULT 4
TMSHP25 2005 bp mRNA linear ROD 03-FEB-1999
LOCUS Tamias asiaticus mRNA for HP-25, complete cds.
DEFINITION D12975
ACCESSION D12975
VERSION D12975.1 GI:287469
KEYWORDS HP-25; collagen-like domain; hibernation-related protein; plasma
protein.
SOURCE Tamias asiaticus liver cDNA to mRNA, clone:PCM25-3.
ORGANISM Tamias sibiricus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Scuridae; Sciurinae;
Tamias.
REFERENCE
1 (bases 1 to 2005)
Takamatsu,N., Ohba,K., Kondo,J., Kondo,N. and Shiba,T.
Hibernation-associated gene regulation of plasma proteins with a
collagen-like domain in mammalian hibernators
Mol. Cell. Biol. 13 (3), 1516-1521 (1993)
93180798
2 (bases 1 to 2005)
Takamatsu,N.
Direct Submission
Submitted (21-AUG-1992) Nobuhiko Takamatsu, School of Hygienic
Sciences, Kitasato University, Laboratory of Molecular Biology;
1-15-1 Kitasato, Sagami-hara, Kanagawa 228, Japan (tel:0427-78-9408,
Fax:0427-78-9403)
FEATURES
Location/Qualifiers
1..2005
/organism="Tamias sibiricus"
/db_xref="taxon:64680"
/clone="PCM25-3"
/tissue_type="liver"
114..761
/notes="collagen-like domain at nt 231-356"
/codon_start=1
/product="HP-25"
/protein_id="BAA02352.1"
/db_xref="GI:287470"
/translation="MPAQGGALSMGAAGFWILVLSITLADSNNOGNSEPCGPRGP
PGPPGIPGPGAPGALGPPGPGVPGIPGPGPPGDEKSSRPSAFKVLSEPRPE
PFQPIVFKALYNQEGHFNMATGEFSCVLPVYNGFDILRFQSSYKIRMRDGIQVR
EKEAQANDSYKHAMGSVIMALGKDKWLESKLKGTSEKGIITHIVFFGYLLYG"
sig_peptide 114..197
mat_peptide 198..758
polyA_signal /product="HP-25"
polyA_site 1983..1988
BASE COUNT 569 a 472 c 494 g 470 t
ORIGIN
Query Match 21.0%; Score 83; DB 10; Length 2005;
Best Local Similarity 54.9%; Pred. No. 7.6e-14;
Matches 186; Conservative 0; Mismatches 150; Indels 3; Gaps 1;
QY 10 CCAGCACACCCAGGCCCCAGAGAAGTGGGGCTCTCGTGTGCACCGAGTTTACCACAA 69
Db 264 CCAGGCTTCCAGCGCTCTCGGAGCACCTGGTCCACAGGACCCCGAGGTGTGCCAGGA 323
QY 70 TATACAGGAGAAATAAGTGAATGACAAA---ATGCCCTGTCTCTGTATAGAAAGTCA 126
Db 324 ATACGAGGACCAACAGTCTCTCGTGGAGAGCTTGAGAAGTGCTCATCGCGACCAATCT 383
QY 127 GCCTTTACTGTGAAGCTCAGTGGAAAACCTTCCTCTCTCTTTCAGAGCCCATCTTTCACA 186
Db 384 GCCTTTCTGTGAAGCTGAGTGAACGCGGCCCCAGAGGCCCTTCCAGGCCCATGTTGCTTCAAG 443
|||||
```

```
QY 187 GGGGTCCTGTACAATGCCAGAGGAGGATTTAAAGAGGAGGCCATGGGAGTCTTTGCTTTCAGG 246
      ||| ||||| ||| ||| ||| ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 444 GAGGCTCTGTACAATCAGGAGGAGGCCACTTCAACATGGCCACGGAGAGTTCAGCTGTGTC 503
      ||| ||||| ||| ||| ||| ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 247 GTGCTTGGGAATTACTACTCCAGCTTTGATGTTGAGCTGCATCATTCGAAGGTGAATATT 306
      ||| ||||| ||| ||| ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 504 CTCCTGTGGTGTACAAATTTGGCTTTTGACATTCGACTTCAGAGTCTCTGTGAAGATC 563
      ||| ||||| ||| ||| ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 307 TGGCTAATGAGGAACAAATTTTGGCTAATAAGGAAGAA 345
      ||||| ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 564 AGGCTCATGAGGAGTGATCCAGGTCAGAGAAAGAA 602
      ||||| ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 5
LOCUS AB067813s3
DEFINITION Callosciurus caniceps HP-25 pseudogene, exon 3, complete cds.
ACCESSION AB069679
VERSION AB069679.1 GI:18149913
SEGMENTS 3 of 3
KEYWORDS Callosciurus caniceps DNA, clone:lambda TS25G1.
SOURCE Callosciurus caniceps
ORGANISM Callosciurus caniceps
REFERENCE 1
AUTHORS Kojima, M., Shiba, T., Kondo, N. and Takamatsu, N.
TITLE The tree squirrel HP-25 gene is a pseudogene
JOURNAL Eur. J. Biochem. 268 (22), 5997-6002 (2001)
MEDLINE 21579740
REFERENCE 2 (bases 1 to 559)
AUTHORS Takamatsu, N.
TITLE Direct Submission
JOURNAL Submitted (06-AGU-2001) Nobuhiko Takamatsu, Kitasato University,
Department of Biosciences, School of Science; 1-15-1 Kitasato,
Sagamihara, Kanagawa 228-8555, Japan
(E-mail: takamatu@jet.sci.kitasato-u.ac.jp, Tel: 81427789408,
Fax: 81427789408)

FEATURES
    source
        1..559
            /organism="Callosciurus caniceps"
            /db_xref="taxon:64664"
            /clone="lambda TS25G1"
            /join(AB067813.1:1503..1658,AB069678.1:26..185,39..559)
            /gene="HP-25"
            /join(AB067813.1:1503..1658,AB069678.1:26..185,39..496)
            /gene="HP-25"
            /pseudo
            /codon_start=1
            /product="HP-25"
            /number=3
            /pseudo

BASE COUNT 150 a 128 c 141 g 140 t
ORIGIN
Query Match 17.3%; Score 68.6; DB 10; Length 559;
Best Local Similarity 55.5%; Pred. No. 1.5e-09;
Matches 156; Conservative 0; Mismatches 119; Indels 6; Gaps 1;

QY 40 GGGCTCTCTGTGCACCAAGGTTTACCACAATATACAGAGAAATAGTGAATGACAAAA 99
      ||| ||||| ||| ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 38 GGACATCCAGGTATACAGGAGGCGGATGACGAGGCCCTCTCGGAGACATGGAGACG 97
      ||||| ||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| |||
QY 100 TGCCCTGCTGTATAGAAGGTCAGCCTTTACTGTGAAGCTCAGTGGAAACTTCTCT 159
      ||||| ||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| |||
Db 98 TGCCCATCACC-----ACCAAAATCTGCCCTTGGCGTGAAGCTGAGTGGCGTCCCA 151
      ||||| ||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| |||
QY 160 CTTCTTTCAAGCCCATCTTTCACAGGGGTCTCTGTACAATGCCAGAGGGAATTAAG 219
      ||| ||||| ||||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| |||
```

```
Db 152 GGGTCTCTCCAGCCCATCATATTCAGGAGTCTCTGTACAAACGGGAGGAGCACTATAC 211
QY 220 GAGGCCATGGGAGTCTTTGCTTGCAGGGTGCCTGGGAATTACTACTCCAGCTTTGATGT 279
      ||| ||||| ||| ||| ||| ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 212 GTGACCACTGGAGAGTTCAGCTATACCAACCCCGGTGTGTACAATTTTGGCTTTGACATT 271
      ||| ||||| ||| ||| ||| ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 280 GAGCTGCATCATTCGAAGGTGAATATTTGGCTTAATGAGGAA 320
      ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| |||
Db 272 GGACTGTTTCAGAGTCTCTGTGAAGATAAGTCTCATGAGAA 312
      ||||| ||||| ||| ||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| |||

RESULT 6
HS302D9/c
LOCUS HS302D9
DEFINITION Human DNA sequence from clone RPI-302D9 on chromosome 22 Contains
          GSSs, complete sequence.
ACCESSION 282198
VERSION 282198.2 GI:6572207
KEYWORDS HTG.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
REFERENCE 1
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE 1 (bases 1 to 145880)
JOURNAL Submitted (08-DEC-1999) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerequest@sanger.ac.uk
On Dec 13, 1999 this sequence version replaced gi:3164067.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence has been finished according to sequence map criteria
as follows. An attempt is made to resolve all sequencing problems,
such as compressions and repeats, but not necessarily within known
annotated human repeat sequence elements (e.g. Alu). Where the
sequence is ambiguous, there is an annotation using the 'unsure'
feature key.
The following abbreviations are used to associate primary accession
numbers given in the feature table with their source databases:
Em., EMBL; Sw., SWISSPROT; Tr., TrEMBL; Wp., WORMPEP; Information
on the WORMPEP database can be found at
http://www.sanger.ac.uk/Projects/C/elegans/wormpep This sequence
was generated from part of bacterial clone contigs of human
chromosome 22, constructed by the Sanger Centre Chromosome 22
Mapping Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr22
RPI-302D9 is from the library RPCI-1 constructed at the Roswell
Park Cancer Institute by the group of Pieter de Jong. For further
details see http://bacpac.med.buffalo.edu/
VECTOR: pCYPAC2
This sequence is the entire insert of clone RPI-302D9 The true left
end of clone CTA-282F2 is at 69682 in this sequence. The true right
end of clone CTA-41562 is at 55167 in this sequence.

FEATURES
    source
        1..145880
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /chromosome="22"
            /clone="RPI-302D9"
            /clone_lib="RPCI-1"
            /note="MER3 repeat: matches 144..209 of consensus"
            /note="MER3 repeat: matches 1..312 of consensus"
            /note="MER3 repeat: matches 1..144 of consensus"
            /note="MER5A repeat: matches 26..187 of consensus"

    repeat_region
        188..245
    repeat_region
        246..571
    repeat_region
        572..759
    repeat_region
        783..933
```

```
repeat_region 1033..1336
/Note="AluSp repeat: matches 1. .299 of consensus"
repeat_region 1450..1583
/Note="MIR repeat: matches 24. .160 of consensus"
repeat_region 1687..1752
/Note="L2 repeat: matches 2593. .2661 of consensus"
repeat_region 2350..2660
/Note="AluSc repeat: matches 3. .309 of consensus"
repeat_region 2684..2981
/Note="AluSg repeat: matches 2. .300 of consensus"
repeat_region 3233..3343
/Note="MLTIE repeat: matches 116. .136 of consensus"
repeat_region 3344..3652
/Note="AluY repeat: matches 1. .309 of consensus"
repeat_region 3653..3928
/Note="MLTIE repeat: matches 136. .359 of consensus"
repeat_region 3929..4278
/Note="THE1B repeat: matches 3. .364 of consensus"
repeat_region 4279..4485
/Note="MLTIE repeat: matches 359. .568 of consensus"
repeat_region 5073..5176
/Note="52 copies 2 mer ct 78 conserved"
repeat_region 5181..5491
/Note="AluJb repeat: matches 1. .311 of consensus"
repeat_region 6369..6485
/Note="L2 repeat: matches 2579. .2705 of consensus"
repeat_region 6647..6685
/Note="WADE1 repeat: matches 1. .23 of consensus"
repeat_region 6686..6987
/Note="AluSx repeat: matches 1. .302 of consensus"
repeat_region 6988..7036
/Note="WADE1 repeat: matches 23. .77 of consensus"
repeat_region 7482..7754
/Note="AluJb repeat: matches 9. .290 of consensus"
repeat_region 7775..8060
/Note="AluJo repeat: matches 1. .295 of consensus"
repeat_region 8414..8551
/Note="L2 repeat: matches 2553. .2706 of consensus"
repeat_region 8914..9030
/Note="MIR repeat: matches 147. .262 of consensus"
repeat_region 9110..9280
/Note="MIR repeat: matches 91. .262 of consensus"
repeat_region 9283..9412
/Note="MIR repeat: matches 15. .144 of consensus"
repeat_region 9521..9679
/Note="FAM repeat: matches 3. .161 of consensus"
repeat_region 9820..10225
/Note="MSTB repeat: matches 2. .425 of consensus"
misc_feature complement(10179..10678)
/Note="match: GSS: Em:B56592"
misc_feature complement(10204..10728)
/Note="match: GSS: Em:A0701486"
misc_feature complement(10249..10706)
/Note="match: GSS: Em:AQ225495"
repeat_region 10312..10383
/Note="MIR repeat: matches 79. .150 of consensus"
misc_feature 10718..11310
/Note="match: GSS: Em:B14024"
misc_feature 10784..11201
/Note="match: GSS: Em:B43656"
repeat_region 11838..11946
/Note="MIR repeat: matches 20. .137 of consensus"
repeat_region 12174..12445
/Note="L2 repeat: matches 1988. .2275 of consensus"
repeat_region 12444..12642
/Note="MIR repeat: matches 63. .241 of consensus"
misc_feature 13017..13369
/Note="match: STS: Em:G49301"
repeat_region 13331..13397
/Note="MIR repeat: matches 174. .244 of consensus"
repeat_region 13398..13698
/Note="AluSp repeat: matches 1. .302 of consensus"
repeat_region 13699..13810
/Note="MIR repeat: matches 76. .174 of consensus"
13806..13919
/Note="MIR repeat: matches 77. .189 of consensus"
13945..14060
/Note="MIR repeat: matches 24. .142 of consensus"
14061..14367
/Note="AluY repeat: matches 1. .301 of consensus"
14368..14452
/Note="MIR repeat: matches 141. .225 of consensus"
14589..14679
/Note="MIR repeat: matches 173. .262 of consensus"
14597..15201
/Note="match: GSS: Em:AQ553482"
14616..15060
/Note="match: GSS: Em:AQ370601"
14868..15040
/Note="MIR repeat: matches 49. .233 of consensus"
15071..15188
/Note="L2 repeat: matches 2112. .2239 of consensus"
15304..15399
/Note="MLT1B repeat: matches 1. .99 of consensus"
15490..15662
/Note="AluSg1 repeat: matches 2. .114 of consensus"
15669..15727
/Note="MLT1B repeat: matches 119. .178 of consensus"
15728..16027
/Note="AluSc repeat: matches 1. .299 of consensus"
16028..16245
/Note="MLT1B repeat: matches 178. .390 of consensus"
16546..16854
/Note="AluY repeat: matches 1. .300 of consensus"
18296..18323
/Note="MSTA repeat: matches 2. .29 of consensus"
18324..18392
/Note="MER66-internal repeat: matches 4919. .4993 of consensus"
18393..18712
/Note="AluJb repeat: matches 1. .311 of consensus"
18713..19133
/Note="MER66-internal repeat: matches 4548. .4919 of consensus"
complement(18872..19230)
/Note="match: GSS: Em:AQ005063"
19251..19719
/Note="match: GSS: Em:B14179"
19537..20290
/Note="HERVH21 repeat: matches 4657. .5784 of consensus"
20317..20382
/Note="33 copies 2 mer ta 68 conserved"
20513..20666
/Note="77 copies 2 mer tt 70 conserved"
20682..21008
/Note="AluSg1 repeat: matches 1. .306 of consensus"
21239..21553
/Note="HUERS-P3 repeat: matches 4410. .4713 of consensus"
21882..22254
/Note="THE1B repeat: matches 1. .364 of consensus"
22302..22537
/Note="MER66-internal repeat: matches 2186. .2417 of consensus"
22538..22850
/Note="AluSp repeat: matches 1. .313 of consensus"
22851..23801
/Note="MER66-internal repeat: matches 1210. .2186 of consensus"
23905..23989
/Note="MER66-internal repeat: matches 3017. .3102 of consensus"
17.2%; Score 68.2; DB 9; Length 145880;
Best Local Similarity 58.7%; Pred.No.3e-09;
Matches 118; Conservative 0; Mismatches 83; Indels 0; Gaps 0;
```

```
Db 90879 AGATCTGATTTGGTGAAGAACTGATCGAGCCCTCCAGTGCCTTAGCCCATTTGT 90820
Qy 180 CTTTCAGGGGTCCTGTACAACTCCAGAGGATTTAAAGAGGCGCATGGAGTCTTTGC 239
Db 90819 CTTCAAGGAAGCCCTTATATAACTACTCAGTTCCATTTCCATTTCTCTGGTGAAGATTAC 90760
Qy 240 TTGCAGGGTGCCTGGGAATTAATACTACTCCAGCTTTGTATGTGTAGCTGCATCATTTGCAAGGT 299
Db 90759 CTGTACCAATCCCTAGTGTGTACCACTTTGGTTTGAATTTGAGCTCTTTTCAGAGCTGTGC 90700
Qy 300 GAATATTTGGCTAATGAGGAA 320
Db 90699 AAATGTGGTATTATGAGGAA 90679

RESULT 7
LOCUS HS732E4
DEFINITION Human DNA sequence from clone CTA-732E4 on chromosome 22q12.1
ACCESSION AL008722
VERSION AL008722.16 GI:5419644
KEYWORDS HTG.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 90497)
Lloyd,D.
Direct Submission
Submitted (08-DEC-1999) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerequest@sanger.ac.uk
On Jul 8, 1999 this sequence version replaced gi:5304865.
This sequence has been finished according to sequence map criteria
as follows. An attempt is made to resolve all sequencing problems,
such as compressions and repeats, but not necessarily within known
annotated human repeat sequence elements (e.g. Alu). Where the
sequence is ambiguous, there is an annotation using the 'unsure'
feature key.
This sequence was generated from part of bacterial clone contigs of
human chromosome 22, constructed by the Sanger Centre Chromosome 22
Mapping Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr22
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
The following abbreviations are used to associate primary accession
numbers given in the feature table with their source databases:
Em., EMBL; Sw., SWISSPROT; Tr., TrEMBL; Wp., WORMPEP; Information
on the WORMPEP database can be found at
http://www.sanger.ac.uk/projects/C-elegans/wormpep
CTA-732E4 is
from the human BAC library described in U-J. Kim et al. (1996)
Genomics 34, 213-218.
VECTOR: pReloBAC11
This sequence is the entire insert of clone CTA-732E4 The true left
end of clone RP11-541J16 is at 5510 in this sequence. The true
right end of clone CTA-544A11 is at 41939 in this sequence.
FEATURES
source
1..90497
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="22"
/map="q12.1"
/clone="CTA-732E4"
/clone.lib="CIT978SK-A2"
repeat_region 2..631
/note="L1M1 repeat: matches 994..1640 of consensus"
repeat_region 632..757
/note="AluYb8 repeat: matches 1..306 of consensus"
/note="AluYb repeat: matches 1..126 of consensus"
758..1385
/note="L1P repeat: matches 3548..4173 of consensus"
1374..1869
/note="L1P repeat: matches 3221..3716 of consensus"
1952..2076
/note="L1MD2 repeat: matches 5142..5259 of consensus"
2760..3005
/note="L1MD2 repeat: matches 5259..6173 of consensus"
3418..3477
/note="30 copies 2 mer ac 90 conserved"
3423..3478
/note="14 copies 4 mer caca 94 conserved"
3670..3742
/note="MER82 repeat: matches 580..653 of consensus"
3763..4128
/note="MER82 repeat: matches 1..386 of consensus"
3879..4353
/note="match: GSS: Em:AQ559531"
4364..4569
/note="L2 repeat: matches 2535..2750 of consensus"
4612..4930
/note="AluJo repeat: matches 1..307 of consensus"
4934..5220
/note="AluX repeat: matches 1..287 of consensus"
5221..5268
/note="16 copies 3 mer taa 97 conserved"
complement(5266..5277)
/note="match: GSS: Em:AQ028870"
complement(5282..5740)
/note="match: GSS: Em:AQ557814"
5417..5569
/note="MIR repeat: matches 2..167 of consensus"
5645..5708
/note="32 copies 2 mer at 71 conserved"
6469..6726
/note="match: STS: Em:G03694; match: STS: Em:G03694"
7573..7749
/note="MIR repeat: matches 22..207 of consensus"
10322..16495
/note="L1PA2 repeat: matches 7..6146 of consensus"
16265..16787
/note="match: GSS: Em:AQ186300"
16926..16967
/note="21 copies 2 mer aa 76 conserved"
17850..17971
/note="AluJo repeat: matches 19..147 of consensus"
17972..18285
/note="AluSg1 repeat: matches 1..309 of consensus"
18286..18472
/note="AluJo repeat: matches 147..309 of consensus"
19710..20005
/note="AluX repeat: matches 1..296 of consensus"
20789..21262
/note="match: GSS: Em:BL4443"
20811..21058
/note="match: STS: Em:L04555"
21369..21551
/note="MIR repeat: matches 2..187 of consensus"
22494..22697
/note="MIR repeat: matches 35..236 of consensus"
23443..23755
/note="AluX repeat: matches 1..312 of consensus"
24052..24348
/note="AluSg repeat: matches 1..295 of consensus"
25623..25670
/note="24 copies 2 mer ac 77 conserved"
26003..26192
/note="MIR repeat: matches 20..216 of consensus"
27067..27138
/note="Charlie4 repeat: matches 1902..1961 of consensus"
27363..27661
/note="AluYb8 repeat: matches 1..306 of consensus"
```


[illegible]

----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: <http://www-seq.wi.mit.edu>
Contact: sequence_submissions@genome.wi.mit.edu
----- Project information
Center project name: L7820
Center clone name: 78_F_17

----- Summary Statistics
Sequencing vector: M13; M77815; 100% of reads
Chemlstry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 124489 bases at least Q40
Consensus quality: 133382 bases at least Q30
Consensus quality: 137772 bases at least Q20
Insert size: 139000; agarose-fp
Insert size: 141014; sum-of-contigs
Quality coverage: 3.5 in Q20 bases; agarose-fp
Quality coverage: 3.4 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 36 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

* 1 1689: contig of 1689 bp in length
* 1690 1789: gap of 100 bp
* 1790 3019: contig of 1230 bp in length
* 3020 3119: gap of 100 bp
* 3120 4921: contig of 1802 bp in length
* 4922 5021: gap of 100 bp
* 5022 7426: contig of 2405 bp in length
* 7427 7526: gap of 100 bp
* 7527 9699: contig of 2173 bp in length
* 9700 9799: gap of 100 bp
* 9800 11899: contig of 2100 bp in length
* 11900 11999: gap of 100 bp
* 12000 14560: contig of 2561 bp in length
* 14561 14660: gap of 100 bp
* 14661 17033: contig of 2373 bp in length
* 17034 17133: gap of 100 bp
* 17134 18747: contig of 1614 bp in length
* 18748 18847: gap of 100 bp
* 18848 21078: contig of 2231 bp in length
* 21079 21178: gap of 100 bp
* 21179 23095: contig of 1917 bp in length
* 23096 23195: gap of 100 bp
* 23196 25510: contig of 2315 bp in length
* 25511 25610: gap of 100 bp
* 25611 27999: contig of 2389 bp in length
* 28000 28099: gap of 100 bp
* 28100 31388: contig of 3289 bp in length
* 31389 31488: gap of 100 bp
* 31489 34192: contig of 2704 bp in length
* 34193 34292: gap of 100 bp
* 34293 37246: contig of 2954 bp in length
* 37247 37346: gap of 100 bp
* 37347 39922: contig of 2576 bp in length
* 39923 40022: gap of 100 bp
* 40023 43695: contig of 3673 bp in length
* 43696 43795: gap of 100 bp
* 43796 48598: contig of 4803 bp in length
* 48599 48698: gap of 100 bp
* 48699 52567: contig of 3869 bp in length
* 52568 52667: gap of 100 bp
* 52668 55272: contig of 2605 bp in length
* 55273 55372: gap of 100 bp
* 55373 59485: contig of 4113 bp in length
* 59486 59585: gap of 100 bp
* 59586 63994: contig of 4409 bp in length

* 63995 64094: gap of 100 bp
* 64095 69097: contig of 5003 bp in length
* 69098 69197: gap of 100 bp
* 69198 73278: contig of 4081 bp in length
* 73279 73378: gap of 100 bp
* 73379 77729: contig of 4351 bp in length
* 77730 77829: gap of 100 bp
* 77830 82571: contig of 4742 bp in length
* 82572 82671: gap of 100 bp
* 82672 87792: contig of 5121 bp in length
* 87793 87892: gap of 100 bp
* 87893 92294: contig of 4402 bp in length
* 92295 92394: gap of 100 bp
* 92395 97160: contig of 4766 bp in length
* 97161 97260: gap of 100 bp
* 97261 102801: contig of 5541 bp in length
* 102802 102901: gap of 100 bp
* 102902 108682: contig of 5781 bp in length
* 108683 108782: gap of 100 bp
* 108783 115194: contig of 6412 bp in length
* 115195 115294: gap of 100 bp
* 115295 123411: contig of 8117 bp in length
* 123412 123511: gap of 100 bp
* 123512 132644: contig of 9133 bp in length
* 132645 132744: gap of 100 bp
* 132745 144514: contig of 11770 bp in length.

FEATURES

Location/Qualifiers
Source
1. .144514
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="18"
/map="18"
/clone_lib="RPC1-11 Human Male BAC"
1. .1689
/note="assembly_fragment"
1790. .3019
/note="assembly_fragment"
3120. .4921
/note="assembly_fragment"
5022. .7426
/note="assembly_fragment"
7527. .9699
/note="assembly_fragment"
9800. .11899
/note="assembly_fragment"
12000. .14560
/note="assembly_fragment"
14661. .17033
/note="assembly_fragment"
17134. .18747
/note="assembly_fragment"
18848. .21078
/note="assembly_fragment"
21179. .23095
/note="assembly_fragment"
clone_end:T7
vector_side:right"
23196. .25510
/note="assembly_fragment"
25611. .27999
/note="assembly_fragment"
28100. .31388
/note="assembly_fragment"
31489. .34192
/note="assembly_fragment"
34293. .37246
/note="assembly_fragment"
37347. .39922
/note="assembly_fragment"
40023. .43695
/note="assembly_fragment"
43796. .48598

```

/note="assembly_fragment"
48699. .52567
/note="assembly_fragment"
52668. .5272
/note="assembly_fragment"
clone_end:Sp6
vector_side:left
55373. .59485
/note="assembly_fragment"
59586. .63994
/note="assembly_fragment"

```

Query Match 16.0%; Score 63.2; DB 2; Length 144514;

Best Local Similarity 74.1%; Pred. No. 9.7e-08;

Matches 80; Conservative 0; Mismatches 28; Indels 0; Gaps 0;

```

QY 289 CATTGCAAGGTGATATTTGGCTAATGAGGAGCAAAATTTGGCTAATGAGGAGAAATTT 348
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 133169 CCTAGAGATTGTGACCTTTGAACCTGAAAGAGATGATTTAGATGATATCAGGGAAGAAATTT 133228
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 349 TCTAAGCAGCAAGATTCACAGAGGTGACTTGGTGCTGCTTAAAGGCA 396
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 133229 TCTAAGCAGCAAGATTCACAGAGGTGACTTGGTGCTGCTTAAAGGCA 133276
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

RESULT 11

```

AC025112 149038 bp DNA linear HTG 08-APR-2000
LOCUS Homo sapiens chromosome 6 clone RP11-608N7 map 6, WORKING DRAFT
DEFINITION SEQUENCE, 27 unordered pieces.
AC025112
VERSION HTG; HTGS_PHASE1; HTGS_DRAFT.
KEYWORDS Homo sapiens.
SOURCE Homo sapiens

```

ORGANISM

```

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 149038)
Birren,B., Linton,L., Nusbaum,C. and Lander,E.
Homo sapiens chromosome 6, clone RP11-608N7
Unpublished
2 (bases 1 to 149038)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
Boquslavsky,L., Boukhgalter,B., Brown,A., Burkett,G.,
Campopiano,A., Castile,A., Choepel,Y., Colangelo,M., Collins,S.,
Collimore,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S.,
Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
Klein,J., LaRoque,K., Lamazares,R., Landers,T., Lehoczyk,J.,
Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
McLirim,J., Meneus,L., Mirhova,T., Miranda,C., Mlenga,V., Morrow,J.,
Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
O'Neill,D., Olivari,T.M., Oliver,J., Peterson,K., Pierre,N.,
Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Testaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
Young,G., Zalnoun,J., Zimmer,A. and Zody,M.

```

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

Web site: <http://www-seq.wi.mit.edu>
 Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information

Center project name: L6882

Center clone name: 608_N7

----- Summary Statistics

Sequencing vector: M13; M77815; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.960731

Consensus quality: 13285 bases at least Q40

Consensus quality: 140748 bases at least Q30

Consensus quality: 143894 bases at least Q20

Insert size: 138000; agarose-fp

Insert size: 146438; sum-of-contigs

Quality coverage: 3.6 in Q20 bases; agarose-fp

Quality coverage: 3.4 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
 * consists of 27 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

```

1 165: contig of 165 bp in length
166 265: gap of 100 bp
266 1355: contig of 1050 bp in length
1356 1455: gap of 100 bp
1456 1752: contig of 297 bp in length
1753 1852: gap of 100 bp
1853 3336: contig of 1484 bp in length
3337 3436: gap of 100 bp
3437 4814: contig of 1378 bp in length
4815 4914: gap of 100 bp
4915 6853: contig of 1939 bp in length
6854 6953: gap of 100 bp
6954 9114: contig of 2161 bp in length
9115 9214: gap of 100 bp
9215 12071: contig of 2857 bp in length
12072 12171: gap of 100 bp
12172 14633: contig of 2462 bp in length
14634 14733: gap of 100 bp
14734 17354: contig of 2621 bp in length
17355 17454: gap of 100 bp
17455 21910: contig of 4456 bp in length
21911 22010: gap of 100 bp
22011 26287: contig of 4277 bp in length
26288 26387: gap of 100 bp
26388 31094: contig of 4707 bp in length
31095 31194: gap of 100 bp
31195 35408: contig of 4214 bp in length
35409 35508: gap of 100 bp
35509 40764: contig of 5256 bp in length
40765 40864: gap of 100 bp
40865 45639: contig of 4775 bp in length
45640 45739: gap of 100 bp
45740 50826: contig of 5087 bp in length
50827 50926: gap of 100 bp
50927 56617: contig of 5691 bp in length
56618 56717: gap of 100 bp
56718 63406: contig of 6689 bp in length
63407 63506: gap of 100 bp
63507 71321: contig of 7815 bp in length
71322 71421: gap of 100 bp
71422 81641: contig of 10220 bp in length
81642 81741: gap of 100 bp
81742 91982: contig of 10241 bp in length
91983 92082: gap of 100 bp
92083 102357: contig of 10275 bp in length
102358 102457: gap of 100 bp
102458 112952: contig of 10495 bp in length
112953 113052: gap of 100 bp

```

Submitted (05-MAR-2000) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 On Apr 8, 2000 this sequence version replaced gi:7158923.
 All repeats were identified using RepeatMasker:

Smit, A.F.A. & Green, P. (1996-1997)

<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

* 113053 124237: contig of 11185 bp in length
* 124238 124337: gap of 100 bp
* 124338 135694: contig of 11357 bp in length
* 135695 135794: gap of 100 bp
* 135795 149038: contig of 13244 bp in length.

FEATURES

```
source
  1. .149038
    /organism="Homo sapiens"
    /db_xref="taxon:9606"
    /chromosome="6"
    /map="6"
    /clone="RP11-608N7"
    /clone_lib="RPCI-11 Human Male BAC"
  1. .165
    /note="assembly_fragment"
    clone_end:T7
    vector_side:right"
misc_feature
  266. .1355
    /note="assembly_fragment"
  1456. .1752
    /note="assembly_fragment"
    clone_end:SP6
    vector_side:left"
misc_feature
  1853. .3336
    /note="assembly_fragment"
  3437. .4814
    /note="assembly_fragment"
  4915. .6853
    /note="assembly_fragment"
  6954. .9114
    /note="assembly_fragment"
  9215. .12071
    /note="assembly_fragment"
  12172. .14633
    /note="assembly_fragment"
  14734. .17354
    /note="assembly_fragment"
  17455. .21910
    /note="assembly_fragment"
  22011. .26287
    /note="assembly_fragment"
  26388. .31094
    /note="assembly_fragment"
  31195. .35408
    /note="assembly_fragment"
  35509. .40764
    /note="assembly_fragment"
  40865. .45639
    /note="assembly_fragment"
  45740. .50826
    /note="assembly_fragment"
  50927. .56617
    /note="assembly_fragment"
  56718. .63406
    /note="assembly_fragment"
  63507. .71321
    /note="assembly_fragment"
  71422. .81641
    /note="assembly_fragment"
  81742. .91982
    /note="assembly_fragment"
  92083. .102357
    /note="assembly_fragment"
  102458. .112952
    /note="assembly_fragment"
  113053. .124237
    /note="assembly_fragment"
  124338. .135694
    /note="assembly_fragment"
  135795. .149038
    /note="assembly_fragment"
  44765 a 26643 c 27220 g 47805 t 2605 others
ORIGIN
```

```
Query Match          16.0%; Score 63.2; DB 2; Length 149038;
Best Local Similarity 74.1%; Pred. No. 9.7e-08;
Matches 80; Conservative 0; Mismatches 28; Indels 0; Gaps 0;

QY 289 CATTGCAAGGTGAATATTGGCTAATGAGGAAGCAAAATTTGGTGAATAGGAAGAATT 348
Db 90741 CCTAGAGATTTGTGAACCTTGAACCTGAAAGAGATGATGATAGATCATCAGGGAAGAATT 90800
|||||
QY 349 TCTAAGCAGCAAAAGCATTCAGAGGTGACTTGGGTGCTGTAAAGGCA 396
Db 90801 TCTAAGCAGCAAAAGCATTCAGAGGTGACTTGGGTGCTGTAAAGGCA 90848
|||||
```

RESULT 12

AC015553/C

LOCUS

DEFINITION

Homo sapiens 9 BAC RP11-100N10 (Roswell Park Cancer Institute Human

BAC Library) complete sequence.

AC015553

AC015553.21 GI:13489133

HTG.

Homo sapiens.

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 162063)

Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,

Alsbrooks,S.L., Amaralunge,H.C., Are,J.R., Banks,T., Barbara,J.,

Benton,J., Bimage,K., Blankenburg,K., Bonnin,D., Bouck,J.,

Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,

Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carlton,T.F.,

Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,

Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,

Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,

Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,

Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,

Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,

Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,

Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,

Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,

Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A.,

Hernandez,J., Hernandez,O., Hodgson,A., Hognes,M., Holloway,C.,

Hollins,B., Honsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J.,

Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,

Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,

Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,

Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,

Louiseged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,

Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,

Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P., Meador,M.,

Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,

Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N.,

Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S.,

Oguh,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B.,

Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L.,

Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojubenok,I., Rolfe,M.,

Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shooshtari,N.,

Sisson,I., Sodergren,E., Sonaik,T., Sparks,A., Stanley,H.,

Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K.,

Tang,H., Tansey,J., Taylor,C., Taylor,T., Teifrod,B., Thomas,N.,

Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R.,

Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C.,

Watlington,S., Williams,G., Williamson,A., Wlecczyk,R., Wooden,S.,

Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.

and Gibbs,R.

Direct Submission

Unpublished

REFERENCE

2 (bases 1 to 162063)

Worley,K.C.

Direct Submission

Submitted (17-NOV-1999) Human Genome Sequencing Center, Department

of Molecular and Human Genetics, Baylor College of Medicine, One

REFERENCE AUTHORS TITLE JOURNAL	Baylor Plaza, Houston, TX 77030, USA 3 (bases 1 to 162063) Worley,K.C. Direct Submission Submitted (30-MAR-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA	37628 41530 68252 68273 68274 68288 87792 98412 98413 113818 113821 113849 122621 133372 134847 160961	aggaagggaag(n)aaggaaggaa ccagcaataa(n)caagttagaat gctggaattg(n)tccttatatt catctaggtg(n)tggttccatg atctaggtg(n)tggttccatg tccatgtttg(n)atacatagat tgggtataag(n)caacagaaga acgtaactta(g)ccaaggtccc cgtaacttag(c)caaggtcccc tgatactgaa(a)ccactaaaaa tactgaacc(n)ctaataaaaa aaaaaaaaa(n)gactaaaaa atggaccatg(n)tgctttgtga gtttataaaa(n)caataaacta tcctttactg(n)ggccttatta aaaaaaaat(c)caaaaacttg
REFERENCE AUTHORS TITLE JOURNAL	Worley,K.C. Direct Submission Submitted (12-JUL-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA	113818 113821 113849 122621 133372 134847 160961	aggaagggaag(n)aaggaaggaa ccagcaataa(n)caagttagaat gctggaattg(n)tccttatatt catctaggtg(n)tggttccatg atctaggtg(n)tggttccatg tccatgtttg(n)atacatagat tgggtataag(n)caacagaaga acgtaactta(n)ccaaggtccc cgtaacttan(n)caaggtcccc tgatactgaa(n)ccnctaaaaa tactgaancc(n)ctaataaaaa aaaaaaaaa(n)gactaaaaa atggaccatg(n)tgctttgtga gtttataaaa(n)caataaacta tcctttactg(n)ggccttatta aaaaaaaat(c)caaaaacttg
REFERENCE AUTHORS TITLE JOURNAL	Worley,K.C. Direct Submission Submitted (13-JUL-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA On Mar 30, 2001 this sequence version replaced gi:13487890. INFORMATION: http://www.hgsc.bcm.tmc.edu/ or email gc-help@bcm.tmc.edu	113818 113821 113849 122621 133372 134847 160961	aggaagggaag(n)aaggaaggaa ccagcaataa(n)caagttagaat gctggaattg(n)tccttatatt catctaggtg(n)tggttccatg atctaggtg(n)tggttccatg tccatgtttg(n)atacatagat tgggtataag(n)caacagaaga acgtaactta(g)ccaaggtccc cgtaacttag(c)caaggtcccc tgatactgaa(a)ccactaaaaa tactgaacc(n)ctaataaaaa aaaaaaaaa(n)gactaaaaa atggaccatg(n)tgctttgtga gtttataaaa(n)caataaacta tcctttactg(n)ggccttatta aaaaaaaat(c)caaaaacttg
COMMENT	CLONE LENGTH: This sequence does not necessarily represent the entire insert of this clone. Overlapping regions of clones are only sequenced and submitted once, so the sequence for the remainder of the insert may be found in the record for the adjacent clones. Overlapping clones are noted at the beginning and end of the Features listing. ANNOTATION OF FEATURES: STSS are identified using ePCR (Genome Res. 7:541-550) searches of a local database that includes entries from dbSTS, GDB, and local mapping efforts. Repeats are identified using RepeatMasker (A. Smit and P. Green, unpublished) for Human and Mouse sequences. Genes and Region of sequence similarity are identified by BLAST (Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the EST and cDNA sequences. Genes demonstrate at least two exons flanked by consensus splice sites that maintained sequence continuity across the splice junctions. Sequences that are not identical matches are annotated as similar. SEQUENCING READ COVERAGE:Sequencing is completed to a minimum standard of double strand coverage with a minimum of 2 clones and 2 reads with no ambiguities or 2 chemistries with a minimum of 2 clones and 3 reads with no ambiguities. If the sequence quality for a region does not meet this standard, it will be indicated in the annotation as Low Coverage. QUALITY OF INDIVIDUAL BASES:This sequence meets stringent quality standards - estimated error rate less than 1 per 10,000 bases. Reports of lowest quality individual bases and measures of base quality are listed below. Description of the metrics can be found at URL: http://gc.bcm.tmc.edu:8088/quality.info/genbank.annotation.html . QUALSTAT-REPORT----- ----- Summary Statistics ----- Contig length: 162063 Phrap values in estimate: 161326 Average error rate (BCM-Phrap estimate): 9.86316e-06 Fraction of Phrap values less than 40 : 0.00234308 Number of consensus changing edits: 18 Number of N's in consensus : 0 ----- Consensus changing edits ----- Position Original+Context Edited+Context 10095 caccacatt(t)accctccaaa caccacatt(a)gcctccaaa 10096 acccacttt(a)ctcccaaa acccactta(g)cctcccaaa		

```

Matches 74; Conservative 0; Mismatches 19; Indels 0; Gaps 0;
Oy 304 ATTTGGCTAATGAGGAACAAATTTTGGCTAATAAAGCAAGAAATTTCTAAGCAGCAAGC 363
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 29373 ACTTTGAACATTGAGAGAGATGATTAGGTATGCGCAGAGAAATTTCTAAGCAGCAAGC 29314
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Oy 364 ATTCAGAGGTGACTTGGTGCTGCTTTAAAGCCA 396
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 29313 ATTCAGAGGTGACTTGGTGCTGCTTTAAAGACA 29281

RESULT 13
AL161623/c
LOCUS
DEFINITION
  Homo sapiens chromosome 9 clone RP11-IK21, *** SEQUENCING IN
  PROGRESS ***, 20 unordered pieces.
ACCESSION
  AL161623
VERSION
  AL161623.12 GI:9930848
KEYWORDS
  HTG; HTGS_PHASE1; HTGS_CANCELLED.
SOURCE
  human.
ORGANISM
  Homo sapiens
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
  .McLay, K.
  Direct Submission
  Submitted (10-JUL-2001) Sanger Centre, Hinxton, Cambridgeshire,
  CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
  requests: clonerequests@sanger.ac.uk
  On Aug 27, 2000 this sequence version replaced gi:9863606.
  ----- Genome Center
  Center: Sanger Centre
  Center code: SC
  Web site: http://www.sanger.ac.uk
  Contact: humquerry@sanger.ac.uk
  ----- Project Information
  Center project name: bALK21
  ----- Summary Statistics
  Assembly program: XGAP4; version 4.5
  Sequencing vector: plasmid; L08752; 100% of reads
  Chemistry: Dye-terminator ET-amersham; 24% of reads
  Dye-terminator Big Dye; 75% of reads
  Consensus quality: 164370 bases at least O40
  Consensus quality: 168213 bases at least O30
  Consensus quality: 170527 bases at least O20
  Insert size: 173234; sum-of-contigs
  Insert size: 168172; 4.4% error; agarose-fp
  Quality coverage: 4.13x in O20 bases; sum-of-contigs Quality
  coverage: 4.57x in O20 bases; agarose-fp
  -----
  * NOTE: This is a 'working draft' sequence. It currently
  * consists of 20 contigs. The true order of the pieces
  * is not known and their order in this sequence record is
  * arbitrary. Gaps between the contigs are represented as
  * runs of N, but the exact sizes of the gaps are unknown.
  * This record will be updated with the finished sequence
  * as soon as it is available and the accession number will
  * be preserved.
  *
  * 1
  * 3512 3611: gap of 100 bp
  * 3612 25895: contig of 22284 bp in length
  * 25896 25995: gap of 100 bp
  * 25996 31543: contig of 5548 bp in length
  * 31544 31643: gap of 100 bp
  * 31644 39248: contig of 7605 bp in length
  * 39249 39348: gap of 100 bp
  * 39349 49948: contig of 10600 bp in length
  * 49949 50048: gap of 100 bp
  * 50049 66131: contig of 16083 bp in length
  * 66132 66231: gap of 100 bp
  * 66232 80131: contig of 13900 bp in length
  * 80132 80231: gap of 100 bp
  * 80232 89785: contig of 9554 bp in length

89786 89885: gap of 100 bp
89886 92978: contig of 3093 bp in length
92979 93078: gap of 100 bp
93079 97528: contig of 4450 bp in length
97529 97628: gap of 100 bp
97629 101520: contig of 3892 bp in length
101521 101620: gap of 100 bp
101621 112966: contig of 11346 bp in length
112967 113066: gap of 100 bp
113067 119217: contig of 6151 bp in length
119218 119317: gap of 100 bp
119318 121556: contig of 2239 bp in length
121557 121656: gap of 100 bp
121657 134400: contig of 12744 bp in length
134401 134500: gap of 100 bp
134501 144934: contig of 10434 bp in length
144935 145034: gap of 100 bp
145035 148818: contig of 3784 bp in length
148819 148918: gap of 100 bp
148919 152178: contig of 3260 bp in length
152179 152278: gap of 100 bp
152279 165020: contig of 12742 bp in length
165021 165120: gap of 100 bp
165121 175134: contig of 10014 bp in length.

FEATURES
      Location/Qualifiers
          source
              1..175134
              /organism="Homo sapiens"
              /db_xref="taxon:9606"
              /chromosome="9"
              /clone="RP11-IK21"
              /clone_lib="RPCI-11.1"
              1..3511
              /note="assembly_fragment:00168"
              fragment_chain:1
          misc_feature
              3612..25895
              /note="assembly_fragment:00194"
              fragment_chain:1
          misc_feature
              25996..31543
              /note="assembly_fragment:00852"
              fragment_chain:2
          misc_feature
              31644..39248
              /note="assembly_fragment:00602"
              fragment_chain:2
          misc_feature
              39349..49948
              /note="assembly_fragment:01593"
              fragment_chain:3
          misc_feature
              50049..66131
              /note="assembly_fragment:01966"
              fragment_chain:3
          misc_feature
              66232..80131
              /note="assembly_fragment:02033"
              fragment_chain:4
          misc_feature
              80232..89785
              /note="assembly_fragment:00372"
              fragment_chain:4
          misc_feature
              89886..92978
              /note="assembly_fragment:02305"
              fragment_chain:5
          misc_feature
              93079..97528
              /note="assembly_fragment:00011"
              fragment_chain:5
          misc_feature
              97629..101520
              /note="assembly_fragment:00955"
              fragment_chain:5
          misc_feature
              101621..112966
              /note="assembly_fragment:01248"
              fragment_chain:5
          misc_feature
              113067..119217
              /note="assembly_fragment:01397"
              fragment_chain:5
          misc_feature
              119318..121556
              /note="assembly_fragment:01523"
              fragment_chain:5
          misc_feature
              121657..134400
              /note="assembly_fragment:01568"
              fragment_chain:5
          misc_feature
              134501..144934
              /note="assembly_fragment:01617"
              fragment_chain:5

```

[illegible]

```
mat_peptide      155..709
                  /product="HP-27"
polyA_signal     1369..1374
polyA_site       1385
BASE COUNT      439 a      285 c      311 g      350 t
ORIGIN

Query Match      15.8%; Score 62.4; DB 10; Length 1385;
Best Local Similarity 55.6%; Pred. No. 1.2e-07;
Matches 120; Conservative 0; Mismatches 96; Indels 0; Gaps 0;

QY 105 CTGTCCTGATATAGAAAGGTCAGCCCTTACTGTGAAGCTCAGTGGAAAACCTTCCTCTTCC 164
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 313 CTGCCACAGCAAAAGGAACATCGGCCCTTTCAGTGAAGGCAAAATGAGCTGCCCCACGCTCC 372
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 165 TTTCAGGCCCATCATCTTCACAGGGTCTCTGTACAATGCCAGAGGGATTAAAGGAGGC 224
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 373 CTCCCAGCCCGTGATCTTCAGGAAGCCCTGCGATGACGCTCAGGGACATTTGATCTGGC 432
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 225 CATGGGAGTCTTTGCTTGCAAGGTGCTGGGAATTACTCCAGCTTTTGATGTTGAGCT 284
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 433 CACTGGTGTGTTCCACCTGCCCACTCCCCAGGACTCTACCAGTTTGGATTTCACATTGAAGC 492
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 285 GCATCATTCGAAGGTGAATATTGGCTAATGAGGAA 320
    || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 493 TGTCAGAGGGCTGTGAAGGTGAGCCTCATGAGAAA 528
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
```

Search completed: February 20, 2003, 04:48:41
Job time : 2075.56 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 19, 2003, 23:48:42 ; Search time 262.925 Seconds
(without alignments)
6406.126 Million cell updates/sec

Title: US-09-997-610-1_COPY_53_156
Perfect score: 104
Sequence: 1 gctggacctccagcacacc.....agtgaaatgacaaatgcc 104

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues
Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*

1: em_estba:*

2: em_esthum:*

3: em_estln:*

4: em_estmu:*

5: em_estov:*

6: em_estpl:*

7: em_estro:*

8: em_htc:*

9: gb_est1:*

10: gb_est2:*

11: gb_htc:*

12: gb_est3:*

13: gb_est4:*

14: gb_est5:*

15: em_estfun:*

16: em_estom:*

17: gb_gss:*

18: em_gss_hum:*

19: em_gss_inv:*

20: em_gss_pln:*

21: em_gss_vrt:*

22: em_gss_fun:*

23: em_gss_mam:*

24: em_gss_mus:*

25: em_gss_other:*

26: em_gss_pro:*

27: em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	38.4	36.9	531	10	BE121287
2	38	36.5	483	13	BI322379
3	38	36.5	552	13	BI863711
4	37.4	36.0	889	14	BQ939199
5	36.8	35.4	695	10	BE285930
6	36.8	35.4	776	13	BI554667

c	7	36.2	34.8	360	14	D69730	D69730
	8	36.2	34.8	632	13	BJ136441	BJ136441
	9	36.2	34.8	726	13	BJ108893	BJ108893
	10	36	34.6	461	13	BI323061	BI323061
	11	35.8	34.4	675	12	BG310518	SMOV3MCA
	12	35.4	34.0	287	10	BE029138	BE029138
	13	35.2	33.8	468	10	BE580165	BE580165
	14	35.2	33.8	483	10	BE579244	BE579244
	15	35.2	33.8	814	10	BE286914	BE286914
	16	34.8	33.5	613	10	BE179928	BE179928
	17	34.6	33.3	300	14	D70222	CELK100E2F
	18	34.6	33.3	360	14	D69877	CELK093C1F
	19	34.6	33.3	407	10	BE581103	BE581103
	20	34.6	33.3	420	10	BE579707	BE579707
	21	34.6	33.3	441	10	BE581944	BE581944
	22	34.6	33.3	478	10	BE579460	BE579460
	23	34.6	33.3	491	10	BE579541	BE579541
	24	34.6	33.3	492	10	BE579271	BE579271
	25	34.6	33.3	496	10	BE579572	BE579572
	26	34.6	33.3	525	10	BE675997	BE675997
	27	34.6	33.3	575	10	AW179710	AW179710
	28	34.6	33.3	585	10	AW179840	AW179840
c	29	34.6	33.3	601	9	AU217866	AU217866
	30	34.6	33.3	605	10	AW347957	AW347957
	31	34.6	33.3	624	12	BE758489	BE758489
	32	34.6	33.3	640	10	AW179745	AW179745
c	33	34.6	33.3	645	13	BJ139557	BJ139557
c	34	34.6	33.3	647	9	AU220646	AU220646
	35	34.6	33.3	647	10	AW409404	AW409404
	36	34.6	33.3	647	10	AW409406	AW409406
	37	34.6	33.3	648	10	AW874708	AW874708
	38	34.6	33.3	654	10	AW348036	AW348036
c	39	34.6	33.3	666	13	BJ144812	BJ144812
c	40	34.6	33.3	672	13	BJ138817	BJ138817
c	41	34.6	33.3	672	13	BJ140783	BJ140783
c	42	34.6	33.3	673	13	BJ144467	BJ144467
c	43	34.6	33.3	674	13	BJ146794	BJ146794
c	44	34.6	33.3	693	9	AU217833	AU217833
	45	34.6	33.3	701	10	AW179849	AW179849

ALIGNMENTS

RESULT 1

BE121287/c

LOCUS BE121287 531 bp mRNA linear EST 13-JUN-2000

DEFINITION UI-R-CA0-baw-b-08-0-UI-sl UI-R-CA0 Rattus norvegicus CDNA clone

ACCESSION BE121287

VERSION BE121287.1 GI:8513392

KEYWORDS EST.

SOURCE Rattus norvegicus

ORGANISM Norway rat.

REFERENCE 1 (bases 1 to 531)

AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.

TITLE Normalization and subtraction: two approaches to facilitate gene discovery

JOURNAL Genome Res. 6 (9), 791-806 (1996)

MEDLINE 97044477

COMMENT Contact: Soares, MB

Program for Rat Gene Discovery and Mapping

University of Iowa

451 Eckstein Medical Research Building Iowa City, IA 52242, USA

Tel: 319 335 8250

Fax: 319 335 9565

Email: msoares@blue.weeg.uiowa.edu

The sequence contained an oligo-dr track that was present in the oligonucleotide that was used to prime the synthesis of first strand cDNA and therefore this may represent a bonafide poly A

BE121287

UI-R-CA0-baw-b-08-0-UI-sl UI-R-CA0 Rattus norvegicus CDNA clone

BE121287

EST

GI:8513392

Rattus.

1 (bases 1 to 531)

Bonaldo,M.F., Lennon,G. and Soares,M.B.

Normalization and subtraction: two approaches to facilitate gene discovery

Genome Res. 6 (9), 791-806 (1996)

Contact: Soares, MB

Program for Rat Gene Discovery and Mapping

University of Iowa

451 Eckstein Medical Research Building Iowa City, IA 52242, USA

Tel: 319 335 8250

Fax: 319 335 9565

Email: msoares@blue.weeg.uiowa.edu

The sequence contained an oligo-dr track that was present in the oligonucleotide that was used to prime the synthesis of first strand cDNA and therefore this may represent a bonafide poly A

McCarter (bchlape@watson.wustl.edu & jmcarter@watson.wustl.edu) at Washington University, St. Louis. DNA Sequencing by: Washington University Genome Sequencing Center St. Louis.

Seq primer: -40RP from Gibco
High quality sequence stop: 395.

FEATURES

Location/Qualifiers
1..552
/organism="Parastrongyloides trichosuri"
/db_xref="taxon:131310"
/clone_lib="Parastrongyloides trichosuri FL pAMP1 v1
Chiapelli McCarter"
/dev_stage="Free Living"
/lab_host="DH10B"

/note="Vector: pAMP1 (Gibco); Site_1: NotI; Site_2: SalI; The library was constructed by Brandi Chiapelli and Dr. James McCarter at Washington University, St. Louis. The cDNA was made by using Dynabead oligo-dT priming (Dyna). PCR based library using a modified protocol from the SMART PCR cDNA Synthesis kit from clontech. Directionally cloned into the UDG sites of pAMP1. Nematodes were provided by Dr. Warwick Grant of AgResearch, New Zealand (warwick.grant@agresearch.co.nz)."

173 a 121 c 134 g 124 t

BASE COUNT

ORIGIN

Query Match 36.5%; Score 38; DB 13; Length 552;
Best Local Similarity 62.8%; Pred. No. 3.2;
Matches 59; Conservative 0; Mismatches 35; Indels 0; Gaps 0;

QY 1 GCTGGACCTCCAGCACACCCCGAGGCCCCCAGAGAAGTGGGGCTCTGGTGCACACAGGT 60

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

DB 250 GATGGATTACCAAGGATTCACAGGAATTAAGGAGAAAGTGGTTCCCTCGAGCACCAGGT 309

QY 61 TTACCACATATACAGGAGCAATAGTGAATCA 94

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

DB 310 ATGCCAGGATTGAAGAGGAGAAATAGAGAAATTA 343

RESULT 4

BQ939199

LOCUS

DEFINITION BQ939199 889 bp mRNA linear EST 21-AUG-2002

IMAGE:6395968 5', mRNA sequence.

ACCESSION BQ939199

VERSION BQ939199.1 GI:22354677

KEYWORDS EST.

SOURCE house mouse.

ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 889)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

AUTHORS NIH-MGC http://mgc.nci.nih.gov/

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL Unpublished (1999)

COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov

Tissue Procurement: The Cepko Laboratory

cDNA Library Preparation: Life Technologies, Inc.

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

Clone Sequencing by: Agencourt Bioscience Corporation

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

Plate: LLAM13891 row: h column: 17

High quality sequence stop: 672.

Location/Qualifiers

1..889

/organism="Mus musculus"

/strain="FVB/N"

/db_xref="taxon:10090"

/clone="IMAGE:6395968"

/clone_lib="NCI_CGAP_Co24"

/lab_host="DH10B (T1 phage-resistant)"

FEATURES

source

/note="Organ: colon; Vector: pCMV-SPORT6; Site_1: NotI; Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT. Average insert size 1.6 kb. Constructed by Life Technologies. Note: this is a NCI_CGAP Library."

BASE COUNT 202 a 251 c 292 g 140 t

ORIGIN

Query Match 36.0%; Score 37.4; DB 14; Length 889;

Best Local Similarity 61.5%; Pred. No. 4.6;

Matches 59; Conservative 0; Mismatches 37; Indels 0; Gaps 0;

QY 2 CTGGACCTCCAGCACACCCCGAGGCCCCCAGAGAAGTGGGGCTCTGGTGCACACAGGT 61

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

DB 600 CAGGACAGTCAGGCCCTCCCTGGCTTCCTGGACAGCAGGGGACACCTGGAGTTCAGGGT 659

QY 62 TACCACATATACAGCAGAAATAGTGAATGACAA 97

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

DB 660 TCCAGGTTCCTANAGGTGAATGGGTGTCATGGGAA 695

RESULT 5

BE285930

LOCUS

DEFINITION BE285930 695 bp mRNA linear EST 26-OCT-2000

IMAGE:601097647 F1 NCI_CGAP_Mam5 Mus musculus cDNA clone IMAGE:3496499 5',

mRNA sequence.

ACCESSION BE285930

VERSION BE285930.1 GI:9163895

KEYWORDS EST.

SOURCE house mouse.

ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 695)

AUTHORS NIH-MGC http://mgc.nci.nih.gov/

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL Unpublished (1999)

COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov

Tissue Procurement: Lothar Hennighausen Ph.D., Robin Humphreys

cDNA Library Preparation: Life Technologies, Inc.

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

Plate: LLAM8548 row: m column: 12

High quality sequence start: 5

High quality sequence stop: 553.

Location/Qualifiers

1..695

/organism="Mus musculus"

/strain="C57BL/6J"

/db_xref="taxon:10090"

/clone="IMAGE:3496499"

/clone_lib="NCI_CGAP_Mam5"

/tissue_type="tumor, gross tissue"

/dev_stage="7 months"

/lab_host="DH10B"

/note="Organ: mammary; Vector: pCMV-SPORT6; Site_1: SalI;

Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT.

Library constructed by Life Technologies. Investigators

providing samples: Lothar Hennighausen/Robin Humphreys,

NIH"

BASE COUNT 175 a 194 c 219 g 107 t

ORIGIN

Query Match

Best Local Similarity

Matches

Conservative

0;

Mismatches

37;

Indels

0;

Gaps

0;

QY

2 CTGGACCTCCAGCACACCCCGAGGCCCCCAGAGAAGTGGGGCTCTGGTGCACACAGGT 61

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

DB 256 CAGGACAGTCAGGCCCTCCCTGGCTTCCTCGACAGCAGGGGACACCTGGAGTTCAGGAT 315

```

QY 62 TACCACAATATACAGGAGAATAAGTGAATGACAA 97
      ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 316 TCCAGAGTTCTAAAGGTGAATGGTGTCTATGGAA 351

RESULT 6
BI554667 776 bp mRNA linear EST 05-SEP-2001
LOCUS 603236619f1 NCI_CGAP_Mam3 Mus musculus cDNA clone IMAGE:5289605 5',
DEFINITION mRNA sequence.
ACCESSION BI554667
VERSION BI554667.1 GI:15441981
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 776)
REFERENCE NIH-MGC http://mgc.nci.nih.gov/.
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: Lothar Hennighausen Ph.D., Chu-Xia Deng Ph.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM11732 row: f column: 06
High quality sequence stop: 759.
FEATURES
      source
      1..776
      /organism="Mus musculus"
      /strain="129,C57BL/6J,FVB/N"
      /db_xref="taxon:10090"
      /clone="IMAGE:5289605"
      /clone_lib="NCI_CGAP_Mam3"
      /tissue_type="tumor, gross tissue"
      /dev_stage="10 months"
      /lab_host="DH10B"
      /note="Organ: mammary; Vector: pCMV-SPORT6; Site.1: SalI;
      Site.2: NotI; Cloned unidirectionally. Primer: Oligo dT.
      Library constructed by Life Technologies. Investigators
      providing samples: Lothar Hennighausen/Chu-Xia Deng, NIH
      Reference for transgenic model: Xu et al., Nature Genetics
      22, 37-43 (1999)."
BASE COUNT 175 a 233 c 231 g 137 t
ORIGIN

      Query Match 35.4%; Score 36.8; DB 13; Length 776;
      Best Local Similarity 61.5%; Pred. No. 6.7;
      Matches 59; Conservative 0; Mismatches 37; Indels 0; Gaps 0;

QY 2 CTGAGCTCTCAGCACACCCAGGCCCGCCAGGAAGTGGGGCTCTCTGTGTCACAGGTT 61
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 233 CAGCAGTCAGCCCTCCCTGGCTTCCTGGACAGGAGGACACCTGGAGTTCCAGGT 292

QY 62 TACCACAATATACAGGAGAATAAGTGAATGACAA 97
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 293 TCCAGAGTTCTAAAGGTGAATGGTGTCTATGGAA 328

RESULT 7
D69730 360 bp mRNA linear EST 07-DEC-1995
LOCUS CELK0730C7F Yuji Kohara unpublished cDNA Caenorhabditis elegans cDNA
DEFINITION clone yk73c7 5', mRNA sequence.
ACCESSION D69730
VERSION D69730.1 GI:1105372
KEYWORDS EST.

```

```

SOURCE Caenorhabditis elegans.
ORGANISM Caenorhabditis elegans
Eukaryota; Metazoa; Nematoidea; Chromadorea; Rhabditida; Rhabditoidea
REFERENCE 1 (bases 1 to 360)
AUTHORS Kohara,Y., Mitsuki,H., Nishigaki,A., Motohashi,T., Sugimoto,A. and
Tabara,H.
Toward an expression map of the C.elegans genome
JOURNAL Unpublished (1994)
COMMENT Contact: Yuji Kohara
Genome Biology Lab.
National Institute of Genetics
Yata 1111, Mishima, Shizuoka 411, Japan
Tel: 81-559-81-6854
Fax: 81-559-81-6855
Email: ykohara@lab.nig.ac.jp.
FEATURES
      source
      1..360
      /organism="Caenorhabditis elegans"
      /strain="CB1489 him-8(el489)"
      /db_xref="taxon:6239"
      /clone="yk73c7"
      /clone_lib="Yuji Kohara unpublished cDNA"
      /sex="hermaphrodite, male"
      /tissue_type="whole animal"
      /dev_stage="varied"
BASE COUNT 96 a 116 c 105 g 42 t 1 others
ORIGIN

      Query Match 34.8%; Score 36.2; DB 14; Length 360;
      Best Local Similarity 65.4%; Pred. No. 9.5;
      Matches 53; Conservative 0; Mismatches 28; Indels 0; Gaps 0;

QY 1 GCTGGACCTCCAGCACACCCAGGCCCGCCAGAGAACTGGGGCTCTGTGTCACACGCT 60
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 177 GCTGGACCAACAGGGGCCATCAGGAGCCCGCCAGGACAAAGGACCTTCAGGAGCCCCAGGA 236

QY 61 TTACCACAATATACAGGAGAA 81
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 237 GTCCAGGACAAACCGGAGGA 257

RESULT 8
BJ136441/c
LOCUS BJ136441
DEFINITION Caenorhabditis elegans cDNA library, C. elegans L1 stage
ACCESSION BJ136441
VERSION BJ136441.1 GI:18296598
KEYWORDS EST.
SOURCE Caenorhabditis elegans.
ORGANISM Caenorhabditis elegans
Eukaryota; Metazoa; Nematoidea; Chromadorea; Rhabditida; Rhabditoidea
REFERENCE 1 (bases 1 to 632)
AUTHORS Kohara,Y., Shin-i.T., Thierry-Mieg,J., Thierry-Mieg,D., Suzuki,Y.
and Sugano,S.
A complementary view of the C.elegans genome
JOURNAL Unpublished (2002)
COMMENT Contact: Tadasu Shin-i
Center For Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tshiniegenes.nig.ac.jp.
FEATURES
      source
      1..632
      /organism="Caenorhabditis elegans"
      /strain="N2"
      /db_xref="taxon:6239"
      /clone="yk113b05"
      /clone_lib="unpublished oligo-capped cDNA library, C.

```

```

elegans L1 stage"
/sex="hermaphrodite"
/tissue_type="whole animal"
/dev_stage="L1"
/note="The AD-wmcdna library was generated with poly(A)+
RNA isolated from both hermaphrodite and male N2 worms of
all larval stages, embryos, adults and dauers and the
subsequent generation of cDNAs by poly(A) priming. The
cDNAs were cloned into ppc86"
BASE COUNT      75 a 163 c 214 g 175 t      5 others
ORIGIN

Query Match      34.8%; Score 36.2; DB 13; Length 632;
Best Local Similarity 65.4%; Pred. No. 9.6;
Matches 53; Conservative 0; Mismatches 28; Indels 0; Gaps 0;

Qy 1 GCTGGACCTCCAGCAGCAGCCGCCAGAGAGTGGGGCTCTCTGGTGACACAGGT 60
||||||| ||||| || || ||||| || || ||||| || || |||||
Db 499 GCTGGACCAACAGCAGCATCAGGAGCCCGCCAGGACAAAGGAGCTCCAGGAGCCCGCAGGA 440
||||||| ||||| || || ||||| || || ||||| || || |||||

Qy 61 TTACCACAATATACAGGAGAA 81
||||||| ||||| || || ||||| || || ||||| || || |||||
Db 439 GCCCAGGACAATCCGGAGGA 419

RESULT 9
Bj108893
LOCUS      726 bp mRNA linear EST 23-JAN-2002
DEFINITION Bj108893 unpublished oligo-capped cDNA library, C. elegans L1 stage
Caenorhabditis elegans cDNA clone yk1113b05 5', mRNA sequence.
ACCESSION Bj108893
VERSION    1
KEYWORDS   EST.
SOURCE     Caenorhabditis elegans.
ORGANISM   Caenorhabditis elegans.

REFERENCE   1 (bases 1 to 726)
AUTHORS    Kohara, Y., Shin-I, T., Thierry-Mieg, J., Thierry-Mieg, D., Suzuki, Y.
and Sugano, S.
TITLE      A complementary view of the C.elegans genome
JOURNAL    Unpublished (2002)
COMMENT    Contact: Tadasu Shin-1
Center For Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tshini@genes.nig.ac.jp.
FEATURES   Location/Qualifiers
            1..726
                /organism="Caenorhabditis elegans"
                /strain="N2"
                /db_xref="taxon:6239"
                /clone="yk1113b05"
                /clone_lib="unpublished oligo-capped cDNA library, C.
elegans L1 stage"
                /sex="hermaphrodite"
                /tissue_type="whole animal"
                /dev_stage="L1"
                /note="The AD-wmcdna library was generated with poly(A)+
RNA isolated from both hermaphrodite and male N2 worms of
all larval stages, embryos, adults and dauers and the
subsequent generation of cDNAs by poly(A) priming. The
cDNAs were cloned into ppc86"
BASE COUNT      189 a 219 c 192 g 125 t      1 others
ORIGIN

Query Match      34.8%; Score 36.2; DB 13; Length 726;
Best Local Similarity 65.4%; Pred. No. 9.6;
Matches 53; Conservative 0; Mismatches 28; Indels 0; Gaps 0;

Qy 1 GCTGGACCTCCAGCAGCAGCCGCCAGAGAGTGGGGCTCTCTGGTGACACAGGT 60

```

```

||||||| ||||| || || ||||| || || ||||| || || |||||
Db 548 GCTGGACCAACAGCAGCATCAGGAGCCCGCCAGGACAAAGGAGCTCCAGGAGCCCGCAGGA 607

Qy 61 TTACCACAATATACAGGAGAA 81
||||||| ||||| || || ||||| || || ||||| || || |||||
Db 608 GCCCAGGACAATCCGGAGGA 628

RESULT 10
Bj1323061
LOCUS      461 bp mRNA linear EST 30-JUL-2001
DEFINITION Kf68h01.yl Strongyloides ratti L2 pAMP1 vl Chiapelli McCarter
Strongyloides ratti cDNA 5' similar to TR:Q19079 Q19079 COSMID
EGAP7. [1] ; contains element MSRL repetitive element ; , mRNA
sequence.
ACCESSION Bj1323061
VERSION    1 GI:15002247
KEYWORDS   EST.
SOURCE     Strongyloides ratti.
ORGANISM   Strongyloides ratti.

REFERENCE   1 (bases 1 to 461)
AUTHORS    McCarter, J., Clifton, S., Chiapelli, B., Pape, D., Martin, J., Wylie, T.,
Dante, M., Marra, M., Hillier, L., Kucaba, T., Theising, B., Bowers, Y.,
Gibbons, M., Ritter, E., Bennett, J., Franklin, C., Tsagarelshvili, R.,
Ronko, I., Kennedy, S., Maguire, L., Beck, C., Underwood, K., Steptoe
, M., Allen, M., Person, B., Swaller, T., Harvey, N., Schurk, R., Kohn, S.,
Shin, T., Jackson, Y., Cardenas, W., McCann, R., Waterston, R. and
Wilson, R.
TITLE      The Washington Univ. Nematode EST Project, 1999
JOURNAL    Unpublished (1999)
COMMENT    Contact: McCarter JP
The Washington Univ. Nematode EST Project, 1999
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
The library was constructed by Brandi Chiapelli and Dr. James
McCarter (bchiapell@watson.wustl.edu & jmcarter@watson.wustl.edu) at
Washington University, St. Louis. DNA Sequencing by: Washington
University Genome Sequencing Center St. Louis.
Seq primer: -40RP from Gibco
High quality sequence stop: 392.
FEATURES   Location/Qualifiers
            1..461
                /organism="Strongyloides ratti"
                /db_xref="taxon:34506"
                /clone_lib="Strongyloides ratti L2 pAMP1 vl Chiapelli
McCarter"
                /dev_stage="L2"
                /lab_host="DH10B"
                /note="vector: pAMP1 (Gibco); The library was constructed
by Brandi Chiapelli and Dr. James McCarter at Washington
University, St. Louis. The cDNA was made by using
Dynabead oligo-dT priming (Dyna). PCR based library
using a modified protocol from the SMART PCR cDNA
Synthesis Kit from Clontech. Directionally cloned into the
UDG sites of pAMP1. Nematodes were provided by Dr. Mark
Viney of Bristol, UK."
BASE COUNT      153 a 85 c 94 g 129 t
ORIGIN

Query Match      34.6%; Score 36; DB 13; Length 461;
Best Local Similarity 64.3%; Pred. No. 11;
Matches 54; Conservative 0; Mismatches 30; Indels 0; Gaps 0;

... Qy 4 GGACCTCCAGCACACCCCGCCCGCCAGAGAGTGGGGCTCTCTGGTGACACAGGTTTA 63
||||||| ||||| || || ||||| || || ||||| || || |||||
Db 365 GGACCAACAGGACCTCCAGGAGCCCGCCAGGAGAGATGGAAGACCGAGGACCGAGGTAGA 424
||||||| ||||| || || ||||| || || ||||| || || |||||

Qy 64 CCACAATATACAGGAGAAATAGT 87

```

LOCUS	BE580165	468 bp	mRNA	linear	EST 09-MAY-2001
DEFINITION	kq37n09.yl TBN95TW-SSR Strongyloides stercoralis cDNA 5' similar to TR:P7958 P79758 COLLAGEN ; contains TAR1.t3 TAR1 repetitive element ; mRNA sequence.				
ACCESSION	BE580165				

```

VERSTON  BE580165.1  GI:98311107
SOURCE    Strongyloides stercoralis.
ORGANISM  Eukaryota; Metazoa; Nematoda; Chromadorea; Strongyloidea;
REFERENCE 1 (bases 1 to 468)
AUTHORS   McCarter, J., Clifton, S., Chiapelli, B., Pape, D., Martin, J., Wylie, T.,
          Dante, M., Marra, M., Hillier, L., Kucaba, T., Theising, B., Bowers, Y.,
          Gibbons, M., Ritter, E., Bennett, J., Franklin, C., Tsagaris, V.,
          Ronko, I., Kennedy, S., Maguire, L., Beck, C., Underwood, K., Steptoe
          M., Allen, M., Person, B., Swaller, T., Harvey, N., Schurk, R., Kohn, S.,
          Shin, T., Jackson, Y., Cardenas, M., McCann, R., Waterston, R. and
          Wilson, R.
TITLE     The Washington Univ. Nematode EST Project, 1999
JOURNAL   Unpublished (1999)
COMMENT   Contact: McCarter JP
          The Washington Univ. Nematode EST Project, 1999
          Washington University School of Medicine
          4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
          Tel: 314 286 1800
          Fax: 314 286 1810
          Email: est@watson.wustl.edu
          The library was constructed by Dr. Thomas Nutman and colleagues of
          NIAID, NIH (tnutman@nih.gov). DNA Sequencing by: Washington
          University Genome Sequencing Center St. Louis.
          High quality sequence stop: 325.
FEATURES  Location/Qualifiers
           source
             1..468
             /organism="Strongyloides stercoralis"
             /strain="Rhabditiform larvae obtained from gerbils"
             /db_xref="taxon:6248"
             /clone_lib="TBN95TM-SSR"
             /lab_host="XL-1 Blue MRF" (Stratagene)
             /note="Vector: Lambda Uni-ZAP XR (Stratagene); Site.1:
             EcoRI; Site.2: XhoI; mRNA was purified from 2 x 10E3
             rhabditiform larvae which had been isolated from gerbils
             experimentally infected with larvae originally isolated
             from experimentally infected dogs. cDNA was constructed
             and, using adaptors, was cloned unidirectionally into the
             vector from the EcoRI site to the XhoI site. The library
             has an unamplified titer of 1 x 10E5 pfu/ml and an
             amplified, undiluted titer of 9 x 10E11 pfu/ml. The
             average insert size of the unamplified library is 675 bp
             (range, 100-1700)."
BASE COUNT 157 a 92 c 134 g 85 t
ORIGIN
... Query Match 33.8%; Score 35.2; DB 10; Length 468;
Best Local Similarity 62.5%; Pred. No. 18;
Matches 55; Conservative 0; Mismatches 33; Indels 0; Gaps 0;
QY 4 GGACCTCCAGCACACCCAGGCCCCGAGAGAGTGGGGCTCTCGTGCACACAGTTTA 63
Db 157 GGACCACAGGACAGATGGATCTCCAGGAGAACAGGACCAAGGTGAACAAGTTTA 216
QY 64 CCACATATACAGGAGAAATAGTGAAA 91
Db 217 CAAGGAGATATTGGACCATCAGAGAAA 244
RESULT 14
LOCUS    BE579244
DEFINITION kq24e08.y1 TBN95TM-SSR Strongyloides stercoralis cDNA 5', similar to
          contains element PTR5 repetitive element ;, mRNA sequence.
ACCESSION BE579244
VERSION   BE579244.1
KEYWORDS EST.
SOURCE   Strongyloides stercoralis.
ORGANISM Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
          Panagrolaimoidea; Strongyloidea;
REFERENCE 1 (bases 1 to 483)
AUTHORS   McCarter, J., Clifton, S., Chiapelli, B., Pape, D., Martin, J., Wylie, T.,
          Dante, M., Marra, M., Hillier, L., Kucaba, T., Theising, B., Bowers, Y.,
          Gibbons, M., Ritter, E., Bennett, J., Franklin, C., Tsagaris, V.,
          Ronko, I., Kennedy, S., Maguire, L., Beck, C., Underwood, K., Steptoe
          M., Allen, M., Person, B., Swaller, T., Harvey, N., Schurk, R., Kohn, S.,
          Shin, T., Jackson, Y., Cardenas, M., McCann, R., Waterston, R. and
          Wilson, R.
TITLE     The Washington Univ. Nematode EST Project, 1999
JOURNAL   Unpublished (1999)
COMMENT   Contact: McCarter JP
          The Washington Univ. Nematode EST Project, 1999
          Washington University School of Medicine
          4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
          Tel: 314 286 1800
          Fax: 314 286 1810
          Email: est@watson.wustl.edu
          The library was constructed by Dr. Thomas Nutman and colleagues of
          NIAID, NIH (tnutman@nih.gov). DNA Sequencing by: Washington
          University Genome Sequencing Center St. Louis.
          High quality sequence stop: 432.
FEATURES  Location/Qualifiers
           source
             1..483
             /organism="Strongyloides stercoralis"
             /strain="Rhabditiform larvae obtained from gerbils"
             /db_xref="taxon:6248"
             /clone_lib="TBN95TM-SSR"
             /lab_host="XL-1 Blue MRF" (Stratagene)
             /note="Vector: Lambda Uni-ZAP XR (Stratagene); Site.1:
             EcoRI; Site.2: XhoI; mRNA was purified from 2 x 10E3
             rhabditiform larvae which had been isolated from gerbils
             experimentally infected with larvae originally isolated
             from experimentally infected dogs. cDNA was constructed
             and, using adaptors, was cloned unidirectionally into the
             vector from the EcoRI site to the XhoI site. The library
             has an unamplified titer of 1 x 10E5 pfu/ml and an
             amplified, undiluted titer of 9 x 10E11 pfu/ml. The
             average insert size of the unamplified library is 675 bp
             (range, 100-1700)."
BASE COUNT 175 a 88 c 131 g 89 t
ORIGIN
... Query Match 33.8%; Score 35.2; DB 10; Length 483;
Best Local Similarity 62.5%; Pred. No. 18;
Matches 55; Conservative 0; Mismatches 33; Indels 0; Gaps 0;
QY 4 GGACCTCCAGCACACCCAGGCCCCGAGAGAGTGGGGCTCTCGTGCACACAGTTTA 63
Db 23 GGACCACAGGACAGATGGATCTCCAGGAGAACAGGACCAAGGTGAACAAGTTTA 82
QY 64 CCACATATACAGGAGAAATAGTGAAA 91
Db 83 CAAGGAGATATTGGACCATCATGAGAAA 110
RESULT 15
LOCUS    BE286914
DEFINITION 601092676F1 NCI_CGAP_Mam5 Mus musculus cDNA clone IMAGE:3487170 5',
          mRNA sequence.
ACCESSION BE286914
VERSION   BE286914.1
KEYWORDS EST.
SOURCE   house mouse.
ORGANISM Mus musculus
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 814)
AUTHORS   NIH-MGC http://mgc.nci.nih.gov/
TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
          Email: cgapbs-re@mail.nih.gov

```

Tissue Procurement: Lothar Hennighausen Ph.D., Robin Humphreys
cDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM8524 row: h column: 19
High quality sequence start: 9
High quality sequence stop: 604.
Location/Qualifiers
I. .814
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:3487170"
/clone_lib="NCI_CGAP_Mam5"
/tissue_type="tumor, gross tissue"
/dev_stage="7 months"
/lab_host="DH10B"
/note="Organ: mammary; Vector: pCMV-SPORT6; Site_1: SalI;
Site_2: NotI; Cloned unidirectionally; Primer: Oligo dT.
Library constructed by Life Technologies. Investigators
providing samples: Lothar Hennighausen/Robin Humphreys,
NIH"

BASE COUNT	213 a	213 c	279 g	108 t	1 others
ORIGIN					
Query Match	33.8%; Score 35.2; DB 10; Length 814;				
Best Local Similarity	60.4%; Pred. No. 18;				
Matches	58; Conservative	0; Mismatches	38; Indels	0; Gaps	0;
QY	2	CTGGACCTCCAGCACACCCAGGCCCCAGAGAAGTGGGGCCTCCTGTGTGCACGAGTT	61		
Db	166	CAGGACAGTCAGGCCTCCCTGGGCTTCCTGGACAGCAGGGGACACCTGGAGTTCCAGGT	225		
QY	62	TACCACAATATACAGGAGAAATAGTGAAATGACAA	97		
Db	226	TCCAGGTTCTAAAGGTGAAATGGGCGTCATGGGAA	261		

Search completed: February 20, 2003, 06:14:21
Job time : 264.925 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 19, 2003, 16:09:30 ; Search time 6.86562 Seconds
(without alignments)
4645.518 Million cell updates/sec

Title: US-09-997-610-1_COPY_53_156

Perfect score: 104

Sequence: 1 gctggacctccagcaccc.....agtgaatgacaaaatgcc 104

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued_Patents_NA.*

- 1: /cgn2_6/ptodata/2/ina/5A_COMB.seq.*
- 2: /cgn2_6/ptodata/2/ina/5B_COMB.seq.*
- 3: /cgn2_6/ptodata/2/ina/6A_COMB.seq.*
- 4: /cgn2_6/ptodata/2/ina/6B_COMB.seq.*
- 5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq.*
- 6: /cgn2_6/ptodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	33.2	31.9	1881	4	US-09-029-348-20
2	30.4	29.2	821	4	US-09-342-681C-7
3	30.4	29.2	1176	4	US-09-342-681C-14
4	30.4	29.2	1574	4	US-09-342-681C-1
5	30.4	29.2	3394	1	US-08-159-784-4
6	28.8	27.7	1898	1	US-08-342-411A-1
7	28.4	27.3	1839	1	US-08-383-744-1
8	28.4	27.3	1839	2	US-08-999-336-1
9	28.4	27.3	1839	5	PCT-US96-01427-1
c 10	28.2	27.1	390	4	US-09-134-001C-635
c 11	28.2	27.1	513	4	US-09-134-001C-647
12	28.2	27.1	585	4	US-09-134-001C-726
c 13	28.2	27.1	810	4	US-09-134-001C-624
14	28.2	27.1	3552	4	US-09-134-001C-693
15	28	26.9	1588	6	5510466-3
16	28	26.9	1507	6	5510466-1
17	28	26.9	5102	6	US-08-494-168-1
18	27.6	26.5	1074	2	US-08-627-151A-15
19	27.6	26.5	1404	6	5171840-8
20	27.6	26.5	1404	6	5480796-8
21	27.6	26.5	1486	4	US-08-795-473B-3
22	27.6	26.5	1486	4	US-09-439-856-3
23	27.6	26.5	2061	6	5171840-1
24	27.6	26.5	2061	6	5480796-1
25	27.6	26.5	3319	4	US-08-795-473B-2
26	27.6	26.5	3319	4	US-09-439-856-2
27	27.4	26.3	2543	1	US-08-555-669-11

ALIGNMENTS

RESULT 1

US-09-029-348-20
; Sequence 20, Application US/09029348
; Patent No. 6171827
; GENERAL INFORMATION:
; APPLICANT: THE VICTORIA UNIVERSITY OF MANCHESTER
; TITLE OF INVENTION: NOVEL PROCOLLAGENS
; FILE REFERENCE: D087857PUS LISTING
; CURRENT APPLICATION NUMBER: US/09/029,348
; CURRENT FILING DATE: 1998-05-07
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 20
; LENGTH: 1881
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: SEQUENCE
; OTHER INFORMATION: DERIVED FROM cDNA OF PROCOLLAGENS
US-09-029-348-20

Query Match 31.9%; Score 33.2; DB 4; Length 1881;
Best Local Similarity 61.6%; Pred. No. 0.088;
Matches 53; Conservative 0; Mismatches 33; Indels 0; Gaps 0;

Qy 2 CTGACCTCCAGCACACCCAGGCCCCAGAGAGAGTGGGCCCTCTGGTGACACAGTT 61

Db 581 CTGGATCTCCAGGATACCAAGGACCCCTGGTGAACCTGGGCAAGCTGGCTTCAGGCC 640

Qy 62 TACCACATATACAGGAGAAATAGT 87

Db 641 CTCAGGACCTCTGGTGTATAGT 666

RESULT 2

US-09-342-681C-7
; Sequence 7, Application US/09342681C
; Patent No. 6355782
; GENERAL INFORMATION:
; APPLICANT: Zonena et al.
; TITLE OF INVENTION: Hypohydrotic ectodermal dysplasia genes and proteins
; FILE REFERENCE: 52978
; CURRENT APPLICATION NUMBER: US/09/342,681C
; CURRENT FILING DATE: 1999-06-29
; PRIOR APPLICATION NUMBER: 60/092,279
; PRIOR FILING DATE: 1998-07-09
; PRIOR APPLICATION NUMBER: 60/112,366
; PRIOR FILING DATE: 1998-12-15
; NUMBER OF SEQ ID NOS: 123
; SOFTWARE: PatentIn Ver. 2.1

```
; SEQ ID NO 7
; LENGTH: 821
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-342-681C-7

Query Match      29.2%; Score 30.4; DB 4; Length 821;
Best Local Similarity 61.2%; Pred. No. 0.54;
Matches 49; Conservative 0; Mismatches 31; Indels 0; Gaps 0;

QY 3 TGGACCTCCAGCACACCCCGCCAGGAGAAAGTGGGGCTCTCTGTGCACCAAGTTT 62
Db 318 TGGCCCTCCAGGACCCCGAGGACCTCCAGGACCCCGAGGAAATTCAGGGAT 377

QY 63 ACCACAATATACAGGAGAAA 82
Db 378 TCCTGGAATTCAGGAACAA 397

RESULT 3
US-09-342-681C-14
; Sequence 14, Application US/09342681C
; Patent No. 6355782
; GENERAL INFORMATION:
; APPLICANT: Zonana et al.
; TITLE OF INVENTION: Hypohydrotic ectodermal dysplasia genes and proteins
; FILE REFERENCE: 52978
; CURRENT APPLICATION NUMBER: US/09/342,681C
; CURRENT FILING DATE: 1999-06-29
; PRIOR APPLICATION NUMBER: 60/092,279
; PRIOR FILING DATE: 1998-07-09
; PRIOR APPLICATION NUMBER: 60/112,366
; PRIOR FILING DATE: 1998-12-15
; NUMBER OF SEQ ID NOS: 123
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 14
; LENGTH: 1176
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1176)
US-09-342-681C-14

Query Match      29.2%; Score 30.4; DB 4; Length 1176;
Best Local Similarity 61.2%; Pred. No. 0.61;
Matches 49; Conservative 0; Mismatches 31; Indels 0; Gaps 0;

QY 3 TGGACCTCCAGCACACCCCGCCAGGAGAAAGTGGGGCTCTCTGTGCACCAAGTTT 62
Db 555 TGGCCCTCCAGGACCCCGAGGACCTCCAGGACCCCGAGGAAATTCAGGGAT 614

QY 63 ACCACAATATACAGGAGAAA 82
Db 615 TCCTGGAATTCAGGAACAA 634

RESULT 4
US-09-342-681C-1
; Sequence 1, Application US/09342681C
; Patent No. 6355782
; GENERAL INFORMATION:
; APPLICANT: Zonana et al.
; TITLE OF INVENTION: Hypohydrotic ectodermal dysplasia genes and proteins
; FILE REFERENCE: 52978
; CURRENT APPLICATION NUMBER: US/09/342,681C
; CURRENT FILING DATE: 1999-06-29
; PRIOR APPLICATION NUMBER: 60/092,279
; PRIOR FILING DATE: 1998-07-09
; PRIOR APPLICATION NUMBER: 60/112,366
; PRIOR FILING DATE: 1998-12-15
; NUMBER OF SEQ ID NOS: 123
; SOFTWARE: PatentIn Ver. 2.1
```

```
; SEQ ID NO 1
; LENGTH: 1574
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (242)..(1417)
US-09-342-681C-1

Query Match      29.2%; Score 30.4; DB 4; Length 1574;
Best Local Similarity 61.2%; Pred. No. 0.68;
Matches 49; Conservative 0; Mismatches 31; Indels 0; Gaps 0;

QY 3 TGGACCTCCAGCACACCCCGCCAGGAGAAAGTGGGGCTCTCTGTGCACCAAGTTT 62
Db 796 TGGCCCTCCAGGACCCCGAGGACCTCCAGGACCCCGAGGAAATTCAGGGAT 855

QY 63 ACCACAATATACAGGAGAAA 82
Db 856 TCCTGGAATTCAGGAACAA 875

RESULT 5
US-08-159-784-4
; Sequence 4, Application US/08159784
; Patent No. 5643783
; GENERAL INFORMATION:
; APPLICANT: Bjorn R. Olsen
; TITLE OF INVENTION: NOVEL COLLAGEN AND USES THEREOF
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; COMPUTER: IBM PS/2 Model 50Z or 55SX
; OPERATING SYSTEM: MS-DOS (Version 5.0)
; SOFTWARE: WordPerfect (Version 5.1)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/159,784
; FILING DATE: December 1, 1993
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLYING APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: John F. Freeman
; REGISTRATION NUMBER: 29,066
; REFERENCE/DOCKET NUMBER: 00246/170001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 542-5070
; TELEFAX: (617) 542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3394
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-159-784-4

Query Match      29.2%; Score 30.4; DB 1; Length 3394;
Best Local Similarity 71.4%; Pred. No. 0.89;
Matches 40; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

QY 4 GGACCTCCAGCACACCCCGCCAGGAGAAAGTGGGGCTCTCTGTGCACCAAGTTT 59
Db 670 GGGCCCTCCAGGACCCCAAGGGGCCCCCAAGAGAAAGTGGGGCCCCCGGACCAAG 725
```

RESULT 6
US-08-342-411A-1
; Sequence 1, Application US/08342411A
; Patent No. 5639616
; GENERAL INFORMATION:
; APPLICANT: LIAO, Shutsung
; APPLICANT: SONG, Ching
; TITLE OF INVENTION: UBIQUITOUS NUCLEAR RECEPTOR:
; TITLE OF INVENTION: COMPOSITIONS AND METHODS
; NUMBER OF SEQUENCES: 38
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: P.O. Box 4433
; CITY: Houston
; STATE: TX
; COUNTRY: USA
; ZIP: 77210-4433
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/342.411A
; FILING DATE: 18-NOV-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: KITCHELL, BARBARA S.
; REGISTRATION NUMBER: 33,928
; REFERENCE/DOCKET NUMBER: ARCD154
; TELEPHONE: (512) 418-3000
; TELEFAX: (713) 789-2679
; TELEX: 79-0924
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1898 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 71..1450
US-08-342-411A-1

Query Match 27.7%; Score 28.8; DB 1; Length 1898;
Best Local Similarity 56.2%; Pred. No. 2.4;
Matches 54; Conservative 0; Mismatches 42; Indels 0; Gaps 0;
Qy 4 GGACCTCCAGCACACCCAGGCCCCAGAGAAGTGGGGCCCTCTCTGGTGACCCAGGTTTA 63
Db 1749 GGACCATGGCTCTCCCTAGCCCGGAGACAGGGGCCCTCTCTCTCTCTCTCTTT 1808
Qy 64 CCACCAATATACAGGAGAAATAGTGAAATGACAAAA 99
Db 1809 ATTTAATAAACTAAAAACAGAAAAA 1844
RESULT 7
US-08-383-744-1
; Sequence 1, Application US/08383744
; Patent No. 5702948
; GENERAL INFORMATION:
; APPLICANT: Greene, Mark I.
; APPLICANT: Davis, James G.
; TITLE OF INVENTION: Saccular collagen and Compositions
; TITLE OF INVENTION: and
; TITLE OF INVENTION: Methods for Making and Using the Same
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz &
; STREET: One Liberty Place, 46th floor
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Wordperfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/999.336
; FILING DATE:
; CLASSIFICATION:

; STREET: One Liberty Place, 46th floor
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Wordperfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/383.744
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: DeLuca, Mark
; REGISTRATION NUMBER: 33,229
; REFERENCE/DOCKET NUMBER: UPN-2039
; TELEPHONE: 215-568-3100
; TELEFAX: 215-568-3439
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1839 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: both
; MOLECULE TYPE: CDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 331..1602
US-08-383-744-1
Query Match 27.3%; Score 28.4; DB 1; Length 1839;
Best Local Similarity 66.1%; Pred. No. 3.2;
Matches 41; Conservative 0; Mismatches 21; Indels 0; Gaps 0;
Qy 28 CCAGAAGAAGTGGGGCCCTCTGGTGACCCAGGTTTACCAATATATACAGAGAAATAGT 87
Db 670 CAAGGACCGGTGGTCTCTCTGGTCCCGATCCAGGAAACACGAGAGAGAAAGCT 729
Qy 88 GA 89
Db 730 GA 731
RESULT 8
US-08-999-336-1
; Sequence 1, Application US/08999336
; Patent No. 5891850
; GENERAL INFORMATION:
; APPLICANT: Greene, Mark I.
; APPLICANT: Davis, James G.
; TITLE OF INVENTION: Saccular collagen and Compositions
; TITLE OF INVENTION: and Methods for Making and Using the Same
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz &
; STREET: No. 5891850ris
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Wordperfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/999.336
; FILING DATE:
; CLASSIFICATION:

```
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/383,744
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: DeLuca, Mark
; REGISTRATION NUMBER: 33,229
; REFERENCE/DOCKET NUMBER: UPN-2039
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-568-3100
; TELEFAX: 215-568-3439
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1839 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: both
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 331..1602
; PCT-US96-01427-1
; US-08-999-336-1

Query Match 27.3%; Score 28.4; DB 2; Length 1839;
Best Local Similarity 66.1%; Pred. No. 3.2;
Matches 41; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

QY 28 CCAGAAGAAGTGGGCTCTCGTGCCACGAGTTTACCACAATATACAGGAGAAATAAGT 87
Db 670 CAAGGACCGGTGGTCTTCTGTGTCCCGAGGAAACACGAGAGAGAAAGT 729

QY 88 GA 89
Db 730 GA 731

RESULT 9
PCT-US96-01427-1
; Sequence 1, Application PC/TUS9601427
; GENERAL INFORMATION:
; APPLICANT: Greene, Mark I.
; APPLICANT: Davis, James G.
; TITLE OF INVENTION: Sacchar collagen and Compositions and
; TITLE OF INVENTION: Methods for Making and Using the Same
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz & Norris
; STREET: One Liberty Place, 46th floor
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Wordperfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US96/01427
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/383,744
; FILING DATE: 02-FEB-1995
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Oeluca, Mark
; REGISTRATION NUMBER: 33,229
; REFERENCE/DOCKET NUMBER: UPN-2653
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-568-3100
; TELEFAX: 215-568-3439
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
```

```
; LENGTH: 1839 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: both
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 331..1602
; PCT-US96-01427-1

Query Match 27.3%; Score 28.4; DB 5; Length 1839;
Best Local Similarity 66.1%; Pred. No. 3.2;
Matches 41; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

QY 28 CCAGAAGAAGTGGGCTCTCGTGCCACGAGTTTACCACAATATACAGGAGAAATAAGT 87
Db 670 CAAGGACCGGTGGTCTTCTGTGTCCCGAGGAAACACGAGAGAGAAAGT 729

QY 88 GA 89
Db 730 GA 731

RESULT 10
US-09-134-001C-635/C
; Sequence 635, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCC
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 635
; LENGTH: 390
; TYPE: DNA
; ORGANISM: Staphylococcus epidermidis
; US-09-134-001C-635

Query Match 27.1%; Score 28.2; DB 4; Length 390;
Best Local Similarity 57.3%; Pred. No. 2.2;
Matches 51; Conservative 0; Mismatches 38; Indels 0; Gaps 0;

QY 2 CTGGACCTCCAGCACACCCCGCCAGGCGCCCGAGAGAAGTGGGCGCTCTGTGTGCACAGGTT 61
Db 161 CAGGTAGCCAGCAGACACCGAGTACGCCAGCAGACAGGTAACCCAGGTAACCCAGGAGGTA 102

QY 62 TACCACAATATACAGGAGAAATAAGTGAA 90
Db 101 CGCCAGCAGAACCCAGGTAAACCCAGCGAA 73

RESULT 11
US-09-134-001C-647/c
; Sequence 647, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCC
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
```

```
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 647
; LENGTH: 513
; TYPE: DNA
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-647

Query Match      27.1%; Score 28.2; DB 4; Length 513;
Best Local Similarity 57.3%; Pred. No. 2.4;
Matches 51; Conservative 0; Mismatches 38; Indels 0; Gaps 0;

Qy 2 CTGGACCTCCAGCACACCCAGGCCCCAGAGAAGTGGGGCCTCTCTGGTGCCACGAGTT 61
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 300 CAGGTACGCCAGCAACAGGTACGCCAGCAGAACCCAGGTAAACCCAGCGGAACCGTA 241
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 62 TACCACAATATACAGGAGAAATAAGTGAA 90
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 240 GCCCAGCAGAACCCAGGTAAACCCAGCGGAA 212

RESULT 12
US-09-134-001C-726
; Sequence 726, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 726
; LENGTH: 585
; TYPE: DNA
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-726

Query Match      27.1%; Score 28.2; DB 4; Length 585;
Best Local Similarity 57.3%; Pred. No. 2.5;
Matches 51; Conservative 0; Mismatches 38; Indels 0; Gaps 0;

Qy 2 CTGGACCTCCAGCACACCCAGGCCCCAGAGAAGTGGGGCCTCTCTGGTGCCACGAGTT 61
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 268 CAGGTACGCCAGCAACAGGTACGCCAGCAGAACCCAGGTAAACCCAGCGGAACCGTA 327
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 62 TACCACAATATACAGGAGAAATAAGTGAA 90
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 328 GCCCAGCAGAACCCAGGTAAACCCAGCGGAA 356

RESULT 13
US-09-134-001C-624/c
; Sequence 624, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 624
; LENGTH: 810
```

```
; TYPE: DNA
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-624

Query Match      27.1%; Score 28.2; DB 4; Length 810;
Best Local Similarity 57.3%; Pred. No. 2.8;
Matches 51; Conservative 0; Mismatches 38; Indels 0; Gaps 0;

Qy 2 CTGGACCTCCAGCACACCCAGGCCCCAGAGAAGTGGGGCCTCTCTGGTGCCACGAGTT 61
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 268 CAGGTACGCCAGCAACAGGTACGCCAGCAGAACCCAGGTAAACCCAGCGGAACCGTA 209
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 62 TACCACAATATACAGGAGAAATAAGTGAA 90
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 208 GCCCAGCAGAACCCAGGTAAACCCAGCGGAA 180

RESULT 14
US-09-134-001C-693
; Sequence 693, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 693
; LENGTH: 3552
; TYPE: DNA
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-693

Query Match      27.1%; Score 28.2; DB 4; Length 3552;
Best Local Similarity 57.3%; Pred. No. 4.7;
Matches 51; Conservative 0; Mismatches 38; Indels 0; Gaps 0;

Qy 2 CTGGACCTCCAGCACACCCAGGCCCCAGAGAAGTGGGGCCTCTCTGGTGCCACGAGTT 61
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3164 CAGGTACGCCAGCAACAGGTACGCCAGCAGAACCCAGGTAAACCCAGCGGAACCGTA 3223
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 62 TACCACAATATACAGGAGAAATAAGTGAA 90
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3224 GCCCAGCAGAACCCAGGTAAACCCAGCGGAA 3252

RESULT 15
5510466-3
; Patent No. 5510466
; APPLICANT: KREIGER, MONTY;KODAMA, TATSUHIKO
; TITLE OF INVENTION: SCAVENGER RECEPTOR PROTEIN AND ANTIBODY
; THERETO
; NUMBER OF SEQUENCES: 12
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/307,400
; FILING DATE: 16-SEP-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 997,113
; FILING DATE: 24-DEC-1992
; APPLICATION NUMBER: 391,486
; FILING DATE: 09-AUG-1989
; APPLICATION NUMBER: 272,002
; FILING DATE: 15-NOV-1988
; SEQ ID NO:3
; LENGTH: 1588
5510466-3
```

Query Match	26.9%	Score 28	DB 6	Length 1588
Best Local Similarity	58.3%	Pred. No. 4.1		
Matches 49	Conservative 0	Mismatches 35	Indels 0	Gaps 0
Qy	4	GGACCTCCAGCACACCCAGGCCCCCCAGAGAAAGTGGGGCCTCTCGTGCACACAGGTTTA	63	
Db	817	GGTCCTCTGGACCTCCAGGTCGAAAGGAGATAGAGGCCCTCTCGGACAAAATGGTATA	876	
Qy	54	CCACAATATACAGGAGAAATAGT	87	
Db	877	CCAGGCTTTCAGGCTCTAATAGT	900	

Search completed: February 19, 2003, 22:58:49
Job time : 10.8656 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 19, 2003, 18:01:21 ; Search time 8.36875 Seconds
(without alignments)
6329.459 Million cell updates/sec

Title: US-09-997-610-1_COPY_53_156
Perfect score: 104
Sequence: 1 gctggacctccagcacaccc.....agtgaatgacaaaatgcc 104

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 424239 seqs, 254661826 residues

Total number of hits satisfying chosen parameters: 848478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications_NA:
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:
9: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:
10: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:
11: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:
12: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:
13: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:
14: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	104	100.0	1381	9	US-09-997-610-1
2	104	100.0	1731	9	US-09-997-610-5
3	75	72.1	1377	9	US-09-997-610-3
4	75	72.1	1731	9	US-09-997-610-7
5	33.2	31.9	561	10	US-09-923-779-47
6	33.2	31.9	617	10	US-09-923-779-51
7	33.2	31.9	648	10	US-09-923-779-66
8	33.2	31.9	653	10	US-09-923-779-53
9	33.2	31.9	655	10	US-09-923-779-68
10	33.2	31.9	656	10	US-09-923-779-69
11	33.2	31.9	658	10	US-09-923-779-71
12	33.2	31.9	659	10	US-09-923-779-52
13	33.2	31.9	823	10	US-09-923-779-5
14	33.2	31.9	823	10	US-09-923-779-28
15	32.8	31.5	392	10	US-09-960-352-1786
16	32.8	31.5	415	10	US-09-960-352-9137
17	32.8	31.5	549	10	US-09-923-779-49
18	32.8	31.5	6158	10	US-09-919-497-6
19	32.8	31.5	6158	10	US-09-954-456-762

20	32.4	31.2	886	10	US-09-923-779-4
21	31.2	30.0	1619	9	US-09-764-868-400
22	31	29.8	3226	10	US-09-954-456-725
23	30.4	29.2	821	9	US-09-729-6588-7
24	30.4	29.2	1176	9	US-09-729-6588-14
25	30.4	29.2	1574	9	US-09-729-6588-1
26	30.4	29.2	2866	9	US-09-764-868-48
27	30.4	29.2	3380	10	US-09-799-799-1
28	30.4	29.2	3394	10	US-09-880-107-2178
29	30.2	29.0	88191	10	US-09-799-799-3
30	29.4	28.3	1320	10	US-09-815-242-7615
31	28.8	27.7	1155	10	US-09-833-381-1293
32	28.6	27.5	684973	10	US-09-263-959-1
33	28.4	27.3	572	10	US-09-923-779-56
34	28	26.9	1234	9	US-09-954-531-1366
35	28	26.9	1344	10	US-09-925-299-44
36	28	26.9	1969	10	US-09-789-561-25
37	27.8	26.7	259	10	US-09-765-231A-49
38	27.6	26.5	3477	9	US-09-935-868-25
39	27.6	26.5	3507	9	US-09-935-868-23
40	27.6	26.5	6158	10	US-09-919-457-6
41	27.6	26.5	6158	10	US-09-954-456-762
42	27.6	26.5	32183	10	US-09-764-869-1494
43	27.4	26.3	2823	10	US-09-919-497-7
44	27.4	26.3	3690	12	US-10-044-090-448
45	27.2	26.2	274	10	US-09-923-876-4300

ALIGNMENTS

RESULT 1
US-09-997-610-1
; Sequence 1, Application US/09997610
; Patent No. US20020156244A1
; GENERAL INFORMATION:
; APPLICANT: Fox, Brian
; APPLICANT: Holloway, James L.
; TITLE OF INVENTION: ADIPOCYTE COMPLIMENT RELATED PROTEIN
; FILE OF INVENTION: ZACR13
; FILE REFERENCE: 00-96
; CURRENT APPLICATION NUMBER: US/09/997,610
; PRIOR FILING DATE: 2001-11-29
; PRIOR APPLICATION NUMBER: US 60/253,924
; PRIOR FILING DATE: 2000-11-29
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: FastSeq for Windows version 3.0
; SEQ ID NO 1
; LENGTH: 1381
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (2)...(1381)
; US-09-997-610-1

Query Match	100.0%	Score 104;	DB 9;	Length 1381;
Best Local Similarity	100.0%	Pred. No. 1.4e-24;		
Matches 104;	Conservative	0;	Mismatches	0;
			Indels	0;
			Gaps	0;
QY	1	GCTGGACCTCCAGCACACCCAGGGCCCCCAGAGAGTGGGGCTCTGGTGCCACGAGT	60	
Db	53	GCTGGACCTCCAGCACACCCAGGGCCCCCAGAGAGTGGGGCTCTGGTGCCACGAGT	112	
QY	61	TTACCACAATATACAGGAGAAATAGTGAATGACAAAATGCC	104	
Db	113	TTACCACAATATACAGGAGAAATAGTGAATGACAAAATGCC	156	
RESULT 2				
US-09-997-610-5				
; Sequence 5, Application US/09997610				
; Patent No. US20020156244A1				

; GENERAL INFORMATION:
; APPLICANT: Fox, Brian
; APPLICANT: Holloway, James L.
; TITLE OF INVENTION: ADIPOCYTE COMPLIMENT RELATED PROTEIN
; TITLE OF INVENTION: ZACRP13
; FILE REFERENCE: 00-96
; CURRENT APPLICATION NUMBER: US/09/997,610
; PRIOR FILING DATE: 2001-11-29
; PRIOR APPLICATION NUMBER: US 60/253,924
; PRIOR FILING DATE: 2000-11-29
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 5
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1731)
US-09-997-610-5

Query Match 100.0%; Score 104; DB 9; Length 1731;
Best Local Similarity 100.0%; Pred. No. 1.5e-24;
Matches 104; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GCTGGACCTCCAGCACACCCAGGCCCCAGAGAAGTGGGGCTCTCTGTGCACCAAGT 60
|||||
Db 406 GCTGGACCTCCAGCACACCCAGGCCCCAGAGAAGTGGGGCTCTCTGTGCACCAAGT 465
QY 61 TTACCAATATACAGGAGAAATAGTGAATGACAAATGCC 104
|||||
Db 466 TTACCAATATACAGGAGAAATAGTGAATGACAAATGCC 509

RESULT 3
US-09-997-610-3
; Sequence 3, Application US/09997610
; Patent No. US20020156244A1
; GENERAL INFORMATION:
; APPLICANT: Fox, Brian
; APPLICANT: Holloway, James L.
; TITLE OF INVENTION: ADIPOCYTE COMPLIMENT RELATED PROTEIN
; TITLE OF INVENTION: ZACRP13
; FILE REFERENCE: 00-96
; CURRENT APPLICATION NUMBER: US/09/997,610
; PRIOR FILING DATE: 2001-11-29
; PRIOR APPLICATION NUMBER: US 60/253,924
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 3
; LENGTH: 1377
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Degenerate polynucleotide encoding a polypeptide
; OTHER INFORMATION: of SEQ ID NO:2
; NAME/KEY: variation
; LOCATION: (1)...(1377)
; OTHER INFORMATION: Each n is independently A, T, G, or C.
; NAME/KEY: misc.feature
; LOCATION: (1)...(1377)
; OTHER INFORMATION: n = A,T,C or G
US-09-997-610-3

Query Match 72.1%; Score 75; DB 9; Length 1377;
Best Local Similarity 64.4%; Pred. No. 3.8e-15;
Matches 67; Conservative 14; Mismatches 23; Indels 0; Gaps 0;
QY 1 GCTGGACCTCCAGCACACCCAGGCCCCAGAGAAGTGGGGCTCTCTGTGCACCAAGT 60
|||||
Db 52 GCNGGNCNCNCNCAYCCNMCCNGCNGARGTNGCNCNGCNCNGCNCNGN 111

QY 61 TTACCAATATACAGGAGAAATAGTGAATGACAAATGCC 104
: || ||||| || ||||| : ||||| ||||| |||||
Db 112 YTNCCNCARTAYACNGGNGARATHWSNGARATGACNAARTGYCC 155
RESULT 4
US-09-997-610-7
; Sequence 7, Application US/09997610
; Patent No. US20020156244A1
; GENERAL INFORMATION:
; APPLICANT: Fox, Brian
; APPLICANT: Holloway, James L.
; TITLE OF INVENTION: ADIPOCYTE COMPLIMENT RELATED PROTEIN
; TITLE OF INVENTION: ZACRP13
; FILE REFERENCE: 00-96
; CURRENT APPLICATION NUMBER: US/09/997,610
; CURRENT FILING DATE: 2001-11-29
; PRIOR APPLICATION NUMBER: US 60/253,924
; PRIOR FILING DATE: 2000-11-29
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 7
; LENGTH: 1731
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Degenerate polynucleotide sequence of zacrp13/zhp1
; OTHER INFORMATION: of SEQ ID NO:6
; NAME/KEY: misc.feature
; LOCATION: (1)...(1731)
; OTHER INFORMATION: n = A,T,C or G
US-09-997-610-7

Query Match 72.1%; Score 75; DB 9; Length 1731;
Best Local Similarity 64.4%; Pred. No. 4.2e-15;
Matches 67; Conservative 14; Mismatches 23; Indels 0; Gaps 0;
QY 1 GCTGGACCTCCAGCACACCCAGGCCCCAGAGAAGTGGGGCTCTCTGTGCACCAAGT 60
|||||
Db 406 GCNGGNCNCNCNCAYCCNMCCNGCNGARGTNGCNCNGCNCNGCNCNGN 465
QY 61 TTACCAATATACAGGAGAAATAGTGAATGACAAATGCC 104
: || ||||| || ||||| : ||||| ||||| |||||
Db 466 YTNCCNCARTAYACNGGNGARATHWSNGARATGACNAARTGYCC 509

RESULT 5
US-09-923-779-47
; Sequence 47, Application US/09923779
; Patent No. US2002007621A1
; GENERAL INFORMATION:
; APPLICANT: Pyle, Ruth A.
; APPLICANT: Xu, Jiangchun
; APPLICANT: Kalos, Michael D.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF PANCREATIC CANCER
; FILE REFERENCE: 210121.553
; CURRENT APPLICATION NUMBER: US/09/923,779
; CURRENT FILING DATE: 2001-08-06
; NUMBER OF SEQ ID NOS: 155
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 47
; LENGTH: 561
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: 544, 550
; OTHER INFORMATION: n = A,T,C or G
US-09-923-779-47
Query Match 31.9%; Score 33.2; DB 10; Length 561;
Best Local Similarity 61.6%; Pred. No. 0.11;


```
Matches 53; Conservative 0; Mismatches 33; Indels 0; Gaps 0;
QY 2 CTGGACCTCCAGCACACCCAGCCGCCAGAGAAGTGGGGCTCTCTGTGCACCAAGTT 61
      ||||| ||||| ||| ||||| ||| ||||| ||||| ||||| |||||
Db 404 CTGGATCTCCAGGATACCAAGAGCCCTGTGTGAACCTGGGCAAGCTGCTCTTCAGGCC 463
QY 62 TACCACAATATACAGGAGAAATAAGT 87
      ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 464 CTCCAGGACCTCTCGTGTATAGGT 489

RESULT 6
US-09-923-779-51
; Sequence 51, Application US/09923779
; Patent No. US20020076721A1
; GENERAL INFORMATION:
; APPLICANT: Pyle, Ruth A.
; APPLICANT: Xu, Jiangchun
; APPLICANT: Kalos, Michael D.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF PANCREATIC CANCER
; FILE REFERENCE: 210121.553
; CURRENT APPLICATION NUMBER: US/09/923,779
; CURRENT FILING DATE: 2001-08-06
; NUMBER OF SEQ ID NOS: 155
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 51
; LENGTH: 617
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 581, 605
; OTHER INFORMATION: n = A,T,C or G
US-09-923-779-51

Query Match 31.9%; Score 33.2; DB 10; Length 617;
Best Local Similarity 61.6%; Pred. No. 0.12;
Matches 53; Conservative 0; Mismatches 33; Indels 0; Gaps 0;
QY 2 CTGGACCTCCAGCACACCCAGCCGCCAGAGAAGTGGGGCTCTCTGTGCACCAAGTT 61
      ||||| ||||| ||| ||||| ||| ||||| ||||| ||||| |||||
Db 404 CTGGATCTCCAGGATACCAAGAGCCCTGTGTGAACCTGGGCAAGCTGCTCTTCAGGCC 463
QY 62 TACCACAATATACAGGAGAAATAAGT 87
      ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 464 CTCCAGGACCTCTCGTGTATAGGT 489

RESULT 7
US-09-923-779-66
; Sequence 66, Application US/09923779
; Patent No. US20020076721A1
; GENERAL INFORMATION:
; APPLICANT: Pyle, Ruth A.
; APPLICANT: Xu, Jiangchun
; APPLICANT: Kalos, Michael D.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF PANCREATIC CANCER
; FILE REFERENCE: 210121.553
; CURRENT APPLICATION NUMBER: US/09/923,779
; CURRENT FILING DATE: 2001-08-06
; NUMBER OF SEQ ID NOS: 155
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 66
; LENGTH: 648
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 642, 646
; OTHER INFORMATION: n = A,T,C or G
US-09-923-779-66

Matches 53; Conservative 0; Mismatches 33; Indels 0; Gaps 0;
QY 2 CTGGACCTCCAGCACACCCAGCCGCCAGAGAAGTGGGGCTCTCTGTGCACCAAGTT 61
      ||||| ||||| ||| ||||| ||| ||||| ||||| ||||| |||||
Db 404 CTGGATCTCCAGGATACCAAGAGCCCTGTGTGAACCTGGGCAAGCTGCTCTTCAGGCC 463
QY 62 TACCACAATATACAGGAGAAATAAGT 87
      ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 464 CTCCAGGACCTCTCGTGTATAGGT 489

RESULT 8
US-09-923-779-53
; Sequence 53, Application US/09923779
; Patent No. US20020076721A1
; GENERAL INFORMATION:
; APPLICANT: Pyle, Ruth A.
; APPLICANT: Xu, Jiangchun
; APPLICANT: Kalos, Michael D.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF PANCREATIC CANCER
; FILE REFERENCE: 210121.553
; CURRENT APPLICATION NUMBER: US/09/923,779
; CURRENT FILING DATE: 2001-08-06
; NUMBER OF SEQ ID NOS: 155
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 53
; LENGTH: 653
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 517, 579, 581, 603, 649
; OTHER INFORMATION: n = A,T,C or G
US-09-923-779-53

Query Match 31.9%; Score 33.2; DB 10; Length 653;
Best Local Similarity 61.6%; Pred. No. 0.12;
Matches 53; Conservative 0; Mismatches 33; Indels 0; Gaps 0;
QY 2 CTGGACCTCCAGCACACCCAGCCGCCAGAGAAGTGGGGCTCTCTGTGCACCAAGTT 61
      ||||| ||||| ||| ||||| ||| ||||| ||||| ||||| |||||
Db 404 CTGGATCTCCAGGATACCAAGAGCCCTGTGTGAACCTGGGCAAGCTGCTCTTCAGGCC 463
QY 62 TACCACAATATACAGGAGAAATAAGT 87
      ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 464 CTCCAGGACCTCTCGTGTATAGGT 489

RESULT 9
US-09-923-779-68
; Sequence 68, Application US/09923779
; Patent No. US20020076721A1
; GENERAL INFORMATION:
; APPLICANT: Pyle, Ruth A.
; APPLICANT: Xu, Jiangchun
; APPLICANT: Kalos, Michael D.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF PANCREATIC CANCER
; FILE REFERENCE: 210121.553
; CURRENT APPLICATION NUMBER: US/09/923,779
; CURRENT FILING DATE: 2001-08-06
; NUMBER OF SEQ ID NOS: 155
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 68
; LENGTH: 655
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
```

```
; LOCATION: 654
; OTHER INFORMATION: n = A,T,C or G
US-09-923-779-68

Query Match          31.9%; Score 33.2; DB 10; Length 655;
Best Local Similarity 61.6%; Pred. No. 0.12;
Matches 53; Conservative 0; Mismatches 33; Indels 0; Gaps 0;

QY 2 CTGGACCTCCAGCACACCCAGGCCGCCAGAGAAGTGGGGCTCTCTGTGCACCAAGTT 61
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 404 CTGGATCTCCAGATACCAAGGACCCCTGGTGAACCTGGGCAAGCTGGTCTTCAGGCC 463

QY 62 TACCACAATATACAGGAGAAATAAGT 87
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 464 CTCAGGACCTCTCTGGTGCTATAGGT 489

RESULT 10
US-09-923-779-69
; Sequence 69, Application US/09923779
; Patent No. US20020076721A1
; GENERAL INFORMATION:
; APPLICANT: Pyle, Ruth A.
; APPLICANT: Xu, Jiangchun
; APPLICANT: Kalos, Michael D.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.553
; CURRENT APPLICATION NUMBER: US/09/923,779
; CURRENT FILING DATE: 2001-08-06
; NUMBER OF SEQ ID NOS: 155
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 69
; LENGTH: 656
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 565, 619, 621
; OTHER INFORMATION: n = A,T,C or G
US-09-923-779-69

Query Match          31.9%; Score 33.2; DB 10; Length 656;
Best Local Similarity 61.6%; Pred. No. 0.12;
Matches 53; Conservative 0; Mismatches 33; Indels 0; Gaps 0;

QY 2 CTGGACCTCCAGCACACCCAGGCCGCCAGAGAAGTGGGGCTCTCTGTGCACCAAGTT 61
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 404 CTGGATCTCCAGATACCAAGGACCCCTGGTGAACCTGGGCAAGCTGGTCTTCAGGCC 463

QY 62 TACCACAATATACAGGAGAAATAAGT 87
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 464 CTCAGGACCTCTCTGGTGCTATAGGT 489

RESULT 11
US-09-923-779-71
; Sequence 71, Application US/09923779
; Patent No. US20020076721A1
; GENERAL INFORMATION:
; APPLICANT: Pyle, Ruth A.
; APPLICANT: Xu, Jiangchun
; APPLICANT: Kalos, Michael D.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.553
; CURRENT APPLICATION NUMBER: US/09/923,779
; CURRENT FILING DATE: 2001-08-06
; NUMBER OF SEQ ID NOS: 155
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 71
; LENGTH: 658
; TYPE: DNA
; ORGANISM: Homo sapiens

; LOCATION: 654
; OTHER INFORMATION: n = A,T,C or G
US-09-923-779-68

Query Match          31.9%; Score 33.2; DB 10; Length 655;
Best Local Similarity 61.6%; Pred. No. 0.12;
Matches 53; Conservative 0; Mismatches 33; Indels 0; Gaps 0;

QY 2 CTGGACCTCCAGCACACCCAGGCCGCCAGAGAAGTGGGGCTCTCTGTGCACCAAGTT 61
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 404 CTGGATCTCCAGATACCAAGGACCCCTGGTGAACCTGGGCAAGCTGGTCTTCAGGCC 463

QY 62 TACCACAATATACAGGAGAAATAAGT 87
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 464 CTCAGGACCTCTCTGGTGCTATAGGT 489

RESULT 12
US-09-923-779-52
; Sequence 52, Application US/09923779
; Patent No. US20020076721A1
; GENERAL INFORMATION:
; APPLICANT: Pyle, Ruth A.
; APPLICANT: Xu, Jiangchun
; APPLICANT: Kalos, Michael D.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.553
; CURRENT APPLICATION NUMBER: US/09/923,779
; CURRENT FILING DATE: 2001-08-06
; NUMBER OF SEQ ID NOS: 155
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 52
; LENGTH: 659
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 627
; OTHER INFORMATION: n = A,T,C or G
US-09-923-779-52

Query Match          31.9%; Score 33.2; DB 10; Length 659;
Best Local Similarity 61.6%; Pred. No. 0.12;
Matches 53; Conservative 0; Mismatches 33; Indels 0; Gaps 0;

QY 2 CTGGACCTCCAGCACACCCAGGCCGCCAGAGAAGTGGGGCTCTCTGTGCACCAAGTT 61
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 404 CTGGATCTCCAGATACCAAGGACCCCTGGTGAACCTGGGCAAGCTGGTCTTCAGGCC 463

QY 62 TACCACAATATACAGGAGAAATAAGT 87
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 464 CTCAGGACCTCTCTGGTGCTATAGGT 489

RESULT 13
US-09-923-779-5/c
; Sequence 5, Application US/09923779
; Patent No. US20020076721A1
; GENERAL INFORMATION:
; APPLICANT: Pyle, Ruth A.
; APPLICANT: Xu, Jiangchun
; APPLICANT: Kalos, Michael D.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.553
; CURRENT APPLICATION NUMBER: US/09/923,779
; CURRENT FILING DATE: 2001-08-06
; NUMBER OF SEQ ID NOS: 155
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 780
; TYPE: DNA
; ORGANISM: Homo sapiens
```

```
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 456, 489, 532, 562, 608, 622, 628, 640, 659, 672, 675, 701,
; LOCATION: 704, 712, 718, 772, 779
; OTHER INFORMATION: n = A,T,C or G
US-09-923-779-5

Query Match      31.9%; Score 33.2; DB 10; Length 780;
Best Local Similarity 61.6%; Pred. No. 0.13;
Matches 53; Conservative 0; Mismatches 33; Indels 0; Gaps 0;

Qy  2 CTGGACCTCCAGCACACCCAGGCCCCAGAGAGTGGGGCTCTCTGGTGACACAGGTT 61
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db  392 CTGGATCTCCAGGATACCAAGGACCCCTGGTGAACCTGGGCAAGCTGGTCTTCAGGCC 333
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy  62 TACCACAATATACAGGAGAAATAAGT 87
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db  332 CTCAGGACCTCCTGGTGCTATAGT 307
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 14
US-09-923-779-28
; Sequence 28, Application US/09923779
; Patent No. US20020076721A1
; GENERAL INFORMATION:
; APPLICANT: Pyle, Ruth A.
; APPLICANT: Xu, Jiangchun
; APPLICANT: Kalos, Michael D.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF PANCREATIC CANCER
; FILE REFERENCE: 210121.553
; CURRENT APPLICATION NUMBER: US/09/923.779
; CURRENT FILING DATE: 2001-08-06
; NUMBER OF SEQ ID NOS: 155
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 28
; LENGTH: 823
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-923-779-28

Query Match      31.9%; Score 33.2; DB 10; Length 823;
Best Local Similarity 61.6%; Pred. No. 0.13;
Matches 53; Conservative 0; Mismatches 33; Indels 0; Gaps 0;

Qy  2 CTGGACCTCCAGCACACCCAGGCCCCAGAGAGTGGGGCTCTCTGGTGACACAGGTT 61
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db  413 CTGGATCTCCAGGATACCAAGGACCCCTGGTGAACCTGGGCAAGCTGGTCTTCAGGCC 472
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy  62 TACCACAATATACAGGAGAAATAAGT 87
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db  473 CTCAGGACCTCCTGGTGCTATAGT 498
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 15
US-09-960-352-1786
; Sequence 1786, Application US/09960352
; Patent No. US20020137139A1
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Tao, Nengbing
; APPLICANT: Byatt, John C.
; APPLICANT: Mathielsen, Nagaippen
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
; FILE REFERENCE: 16511.006/37-21(10298)C
; CURRENT APPLICATION NUMBER: US/09/960.352
; CURRENT FILING DATE: 2001-09-24
; NUMBER OF SEQ ID NOS: 15112
; SEQ ID NO 1786
; LENGTH: 392
; TYPE: DNA
; ORGANISM: Bos taurus
```

```
; OTHER INFORMATION: Clone ID: 08-LIB34-023-Q1-E1-B7
US-09-960-352-1786

Query Match      31.5%; Score 32.8; DB 10; Length 392;
Best Local Similarity 64.5%; Pred. No. 0.13;
Matches 49; Conservative 0; Mismatches 27; Indels 0; Gaps 0;

Qy  2 CTGGACCTCCAGCACACCCAGGCCCCAGAGAGTGGGGCTCTCTGGTGACACAGGTT 61
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db  153 CGGGCCACACAGGACACCCAGGCCCTCCAGGTATTAAGAGGGCTCCAGGTATAAGAGGAA 212
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy  62 TACCACAATATACAGG 77
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db  213 TACCAGGTTTGCCAGG 228
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Search completed: February 19, 2003, 23:02:31
Job time : 11.3688 secs
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 19, 2003, 15:08:07 ; Search time 35.9937 Seconds
(without alignments)
6506.903 Million cell updates/sec

Title: us-09-997-610-1-copy_53_156

Perfect score: 104

Sequence: 1 gctggacctccagcacccc.....agtgaatgacaaaatgccc 104

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries.

Database : N_Geneseq_101002.*

- 1: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*
- 2: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*
- 3: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*
- 4: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*
- 5: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*
- 6: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*
- 7: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*
- 8: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*
- 9: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*
- 10: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*
- 11: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
- 12: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*
- 13: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*
- 14: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*
- 15: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*
- 16: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*
- 17: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*
- 18: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*
- 19: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*
- 20: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*
- 21: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*
- 22: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.*
- 23: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.*
- 24: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	59	56.7	1338	24	Human genset metab
2	36.8	35.4	6512	24	Mouse ischaemic co
3	34	32.7	5467	22	Human EST-derived
4	34	32.7	5468	22	Human EST-derived
5	33.2	31.9	561	24	CDNA #47 encoding
6	33.2	31.9	617	24	CDNA #51 encoding
7	33.2	31.9	648	24	CDNA #66 encoding
8	33.2	31.9	653	24	CDNA #53 encoding
9	33.2	31.9	655	24	CDNA #68 encoding

10	33.2	31.9	656	24	ABK44129	CDNA #69 encoding
11	33.2	31.9	658	24	ABK44131	CDNA #71 encoding
12	33.2	31.9	659	24	ABK44112	CDNA #52 encoding
c	13	33.2	780	24	ABK44065	CDNA #5 encoding h
14	33.2	31.9	823	24	ABK44088	CDNA #28 encoding
15	33.2	31.9	994	22	AAF44902	Human breast cance
16	33.2	31.9	1881	18	AAT59892	Coding sequence fo
17	33.2	31.9	4428	22	AAD06574	Bovine alpha(III)
18	33.2	31.9	4428	22	AAD06575	Bovine alpha(III)
19	33.2	31.9	4428	22	AAD06578	Porcine alpha(III)
20	33.2	31.9	5460	17	AAT16508	Vector PAC3A1 cont
21	33.2	31.9	5460	22	ABA83117	Collagen type III
22	33.2	31.9	5460	24	ABL92101	Human Tumour Endo
23	33.2	31.9	5466	23	AAS79378	DNA encoding novel
24	32.8	31.5	501	13	AAQ26041	Human collagen XI
25	32.8	31.5	549	24	ABK44109	CDNA #49 encoding
c	26	32.8	1121	23	AAS89670	DNA encoding novel
27	32.8	31.5	5010	24	ABL92114	Human Tumour Endo
28	32.8	31.5	6158	24	ABL62095	Colon adenocarcino
29	32.8	31.5	6158	24	ABL65452	Lung cancer relate
30	32.8	31.5	6158	24	ABK35486	Human endometrial
31	32.8	31.5	6436	22	AAI60019	Human polynucleoti
c	32	32.8	51935	22	AAK75883	Human immune/haema
33	32.4	31.2	886	24	ABK44064	CDNA #4 encoding h
34	32.4	31.2	1317	22	AAH48067	Murine HSP47 inter
c	35	32	61710	22	AAK83782	Human immune/haema
36	31.2	30.0	1619	22	AAS27365	CDNA encoding nove
37	31.2	30.0	1619	22	AAS34830	CDNA encoding nove
38	31.2	30.0	6674	22	AAS58233	Human polynucleoti
39	31	29.8	3226	24	ABK46066	Human CDNA differe
40	31	29.8	3226	24	ABL65415	Lung cancer relate
41	30.4	29.2	707	22	AAS44890	Human contig polyn
42	30.4	29.2	821	24	ABL51013	Human EDAL-II exon
43	30.4	29.2	1176	24	ABL51020	Human EDAL-II open
44	30.4	29.2	1574	24	ABL51009	Human EDAL-II enco
45	30.4	29.2	2686	22	AAS27013	CDNA encoding nove

ALIGNMENTS

RESULT 1
ID AAL44066 standard; CDNA; 1338 BP.
XX AC AAL44066;
XX DT 27-SEP-2002 (first entry)
XX DE Human genset metabolic gene (GMG-9) cDNA sequence.
XX DE Human; gene; ss; gene therapy; genset metabolic gene; GMG-7A; GMG-7B;
KW GMG-8; GMG-9; GMG-10; GMG-11; metabolic-related disorder; obesity;
KW impaired glucose tolerance; insulin resistance; Syndrome X;
KW Type II diabetes; hyperlipidaemia; atherosclerosis; hypertension;
KW heart disease; cardiac insufficiency; coronary insufficiency;
KW high blood pressure; insulin sensitizer;
KW non-insulin dependent diabetes mellitus.
XX OS Homo sapiens.

Key	Location/Qualifiers
CDS	1..1338
FT	/*tag= a
FT	/partial
FT	/product= "Human GMG-9 protein"
FT	/note= "No stop codon is given"
PN	WO200255694-A2.
XX	18-JUL-2002.
PD	
XX	
PF	15-JAN-2002; 2002WO-IB01215.

XX 16-JAN-2001; 2001US-262235P.
 XX (GEST) GENSET.
 XX Erickson MR, Bour BA, Bihain B, Tanaka H;
 XX WPI; 2002-557821/59.
 XX P-PSDB; AAO15423.
 XX Treating or preventing a metabolic-related disease or disorder, e.g.
 XX obesity, impaired glucose tolerance, insulin resistance, Syndrome X, or
 XX Type II diabetes, comprises administering Genset Metabolic Genes -
 XX Disclosure; Page 122-124; 128pp; English.
 XX The invention comprises the amino acid and coding sequences of six human
 XX genset metabolic genes (GMG-7A, GMG-7B, GMG-8, GMG-9, GMG-10 and GMG-11).
 XX The GMG DNA and protein sequences of the invention are useful for
 XX treating or preventing metabolic-related disorders, such as: obesity;
 XX impaired glucose tolerance; insulin resistance; Syndrome X; Type II
 XX diabetes; hyperlipidaemia; atherosclerosis; hypertension; and heart
 XX diseases (e.g. cardiac insufficiency, coronary insufficiency or high
 XX blood pressure). The GMG DNA and protein sequences of the invention may
 XX also be used as insulin sensitizers - for improving insulin sensitivity
 XX in persons with non-insulin dependent diabetes mellitus. The present cDNA
 XX sequence encodes the human GMG-9 protein.
 XX Sequence 1338 BP; 359 A; 288 C; 350 G; 341 T; 0 other;
 Query Match 56.7%; Score 59; DB 24; Length 1338;
 Best Local Similarity 92.5%; Pred. No. 2.2e-09;
 Matches 62; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
 QY 38 TGGGCTCTCTGTGACACAGCTTTACCAATATACAGGAGAAATAGTGAATGACAA 97
 Db 50 TGGATGTCTCTGTGCTCCAGGTTTACCAATATACAGGAGAAATAGTGAATGACAA 109
 QY 98 AATGCC 104
 Db 110 AATGCC 116
 RESULT 2
 AB199819
 ID AB199819 standard; cDNA; 6512 BP.
 XX AB199819;
 XX 07-MAR-2002 (first entry)
 XX Mouse ischaemic condition related cDNA sequence SEQ ID NO:932.
 DE Mouse; ischaemia; compressive ischaemia; occlusive ischaemia;
 KW vasospastic ischaemia; ischaemic condition; ischaemic disease; ss.
 XX Mus musculus.
 OS WO200188188-A2.
 PN 22-NOV-2001.
 XX 18-MAY-2001; 2001WO-JP04192.
 XX 18-MAY-2000; 2000JP-0145977.
 XX (UYNI-) UNIV NIHON SCHOOL JURIDICAL PERSON.
 XX Ishikawa K, Asai S, Takahashi Y, Nagata T, Ishii Y;
 XX WPI; 2002-034733/04.
 XX P-PSDB; ABB57334.
 XX

PT Examining the ischemic condition (e.g. occlusive ischemia) by measuring
 PT expression levels of particular genes defined in the specification or
 PT by determining the expression profile of a gene group comprising these
 XX genes -
 XX Claim 2; Page 2340-2352; 2690pp; English.
 XX The present invention describes a method for examining ischaemic
 XX conditions, comprising measuring the expression levels of particular
 XX genes (I) in a test sample or determining the expression profile of a
 XX gene group in the sample comprising genes selected from (I). The method
 XX is useful for examining the ischaemic condition (e.g. compressive
 XX ischaemia, occlusive ischaemia or vasospastic ischaemia) by measuring
 XX expression levels of particular genes (ABI99202 to ABI99912, encoding
 XX the protein sequences in ABB57020 to ABB57374) or by determining the
 XX expression profile of a gene group comprising these genes. The
 XX expression levels or expression profiles produced by these genes are
 XX used as an indicator when screening for ischaemic condition-improving
 XX drugs or therapeutics for ischaemic diseases. ABI99913 and ABI99914
 XX represent PCR primers for a mouse ischaemic condition related sequence,
 XX which are used in the exemplification of the present invention.
 XX Sequence 6512 BP; 1553 A; 1798 C; 1950 G; 1211 T; 0 other;
 Query Match 35.4%; Score 36.8; DB 24; Length 6512;
 Best Local Similarity 61.5%; Pred. No. 0.065;
 Matches 59; Conservative 0; Mismatches 37; Indels 0; Gaps 0;
 QY 2 CTGGACCTCCAGCACACCCAGGCCCCAGAGAGAGTGGGGCTCTCTGTGCACAGGTT 61
 Db 2692 CAGGACAGTCAGGCTCCCTGGCCCTCTCTGGACAGGAGACCTGGAGTTCAGGTT 2751
 QY 62 TACCACAATATACAGGAGAAATAGTGAATGACAA 97
 Db 2752 TCCAGGTTCTAAAGTGAATGGTGTCATGGGAA 2787
 RESULT 3
 AAH98343
 ID AAH98343 standard; cDNA; 5467 BP.
 XX AAH98343;
 XX 12-OCT-2001 (first entry)
 XX Human EST-derived coding sequence SEQ ID NO: 200.
 DE Human
 KW Human; sheep; pig; cow; fruit fly; yeast; hamster; macaque; horse;
 KW tomato; monkey; dog; sea urchin; expressed sequence tag; EST;
 KW diagnostics; forensic test; gene mapping; genetic disorder;
 KW biodiversity; gene therapy; nutrition; ss.
 XX Homo sapiens.
 OS WO200154477-A2.
 PN 02-AUG-2001.
 XX 25-JAN-2001; 2001WO-US02687.
 XX 25-JAN-2000; 2000US-0491404.
 XX 17-JUL-2000; 2000US-0617746.
 XX 03-AUG-2000; 2000US-0631451.
 XX 15-SEP-2000; 2000US-0663870.
 XX (HYSE-) HYSEQ INC.
 XX Tang YT, Liu C, Zhou P, Qian XB, Wang Z, Chen R, Asundi V;
 XX Cao Y, Drmanac RA, Zhang J, Werhman T;
 XX WPI; 2001-476164/51.
 XX P-PSDB; AAM23684.
 XX

PT Isolated polypeptide for treatment of diseases, diagnostics, raising
 PT antibodies and research use -
 XX
 PS Claim 1: Page 314-315; 1275pp; English.
 XX
 CC The present invention provides the protein and coding sequences of novel
 CC proteins from a variety of organisms, including human, dog, cat, horse,
 CC cow, pig, hamster, monkey, macaque, yeast, bacteria, fruit fly, sea
 CC urchin and tomato. These were derived from expressed sequence tags (ESTs)
 CC from the organism of interest. They can be used in diagnostics,
 CC forensics, gene mapping, identification of mutations, to assess
 CC biodiversity and for nutritional purposes. The present sequence is a cDNA
 CC of the invention.
 XX
 SQ Sequence 5467 BP; 1330 A; 1323 C; 1550 G; 1264 T; 0 other;
 Query Match 32.7%; Score 34; DB 22; Length 5467;
 Best Local Similarity 61.1%; Pred. No. 0.5;
 Matches 55; Conservative 0; Mismatches 35; Indels 0; Gaps 0;
 QY 1 GCTGGACCTCCAGCACACCCAGGCCGCCAGAGAGTGGGCCCTCTGGTGCACACAGGT 60
 Db 631 GCTGGCCCCCAGGCCGCCGCCGCCCTGTACATCTGTGTCATCTGTGTCCTGTTCCCTCGA 690
 QY 61 TTACCACATATACAGGAGCAATAAGTGAA 90
 Db 691 TCTCCAGGATACCAAGGAGGCCCTGTGTGAA 720
 RESULT 4
 AAH98411
 ID AAH98411 standard; cDNA; 5468 BP.
 XX
 AC AAH98411;
 XX
 DT 12-OCT-2001 (first entry)
 XX
 DE Human EST-derived coding sequence SEQ ID NO: 268.
 XX
 KW Human: sheep; pig; cow; fruit fly; yeast; hamster; macaque; horse;
 KW tomato; monkey; dog; sea urchin; expressed sequence tag; EST;
 KW diagnostics; forensic test; gene mapping; genetic disorder;
 KW biodiversity; gene therapy; nutrition; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO200154477-A2.
 XX
 PD 02-AUG-2001.
 XX
 PF 25-JAN-2001; 2001WO-US02687.
 XX
 PR 25-JAN-2000; 2000US-0491404.
 PR 17-JUL-2000; 2000US-0617746.
 PR 03-AUG-2000; 2000US-0631451.
 PR 15-SEP-2000; 2000US-0663870.
 XX
 PA (HYSE-) HYSEQ INC.
 XX
 PI Tang YT, Liu C, Zhou P, Qian XB, Wang Z, Chen R, Asundi V;
 PI Cao Y, Drmanac RA, Zhang J, Werhman T;
 DR
 DR WPI; 2001-476164/51.
 DR P-PSDB; AAM23752.
 XX
 XX Isolated polypeptide for treatment of diseases, diagnostics, raising
 PT antibodies and research use -
 XX
 PS Claim 1: Page 381-382; 1275pp; English.
 XX
 CC The present invention provides the protein and coding sequences of novel
 CC proteins from a variety of organisms, including human, dog, cat, horse,
 CC cow, pig, hamster, monkey, macaque, yeast, bacteria, fruit fly, sea

CC urchin and tomato. These were derived from expressed sequence tags (ESTs)
 CC from the organism of interest. They can be used in diagnostics,
 CC forensics, gene mapping, identification of mutations, to assess
 CC biodiversity and for nutritional purposes. The present sequence is a cDNA
 CC of the invention.
 XX
 SQ Sequence 5468 BP; 1330 A; 1323 C; 1551 G; 1264 T; 0 other;
 Query Match 32.7%; Score 34; DB 22; Length 5468;
 Best Local Similarity 61.1%; Pred. No. 0.5;
 Matches 55; Conservative 0; Mismatches 35; Indels 0; Gaps 0;
 QY 1 GCTGGACCTCCAGCACACCCAGGCCGCCAGAGAGTGGGCCCTCTGGTGCACACAGGT 60
 Db 631 GCTGGCCCCCAGGCCGCCGCCGCCCTGTACATCTGTGTCATCTGTGTCCTGTTCCCTCGA 690
 QY 61 TTACCACATATACAGGAGCAATAAGTGAA 90
 Db 691 TCTCCAGGATACCAAGGAGGCCCTGTGTGAA 720
 RESULT 5
 ABK44107
 ID ABK44107 standard; cDNA; 561 BP.
 XX
 AC ABK44107;
 XX
 DT 21-MAY-2002 (first entry)
 XX
 DE cDNA #47 encoding human pancreatic tumour protein.
 XX
 KW Human; pancreatic tumour protein; immune response; pancreatic cancer;
 KW development of cancer; cancer progression; cytostatic; gene; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO200212331-A2.
 XX
 PD 14-FEB-2002.
 XX
 PF 06-AUG-2001; 2001WO-US24619.
 XX
 PR 07-AUG-2000; 2000US-223130P.
 PR 30-JAN-2001; 2001US-265447P.
 PR 15-MAY-2001; 2001US-291201P.
 XX
 PA (CORI-) CORIXA CORP.
 XX
 PI Pyle RA, Xu J, Kalos MD;
 XX
 DR WPI; 2002-241741/29.
 XX
 PT Novel polynucleotide encoding pancreatic tumour polypeptides, useful in
 PT pharmaceutical compositions, e.g. vaccines, for treating pancreatic
 PT cancers -
 XX
 PS Claim 1; Page 126; 167pp; English.
 XX
 CC The present invention relates to the isolation of cDNA sequences
 CC encoding human pancreatic tumour proteins. The polynucleotide
 CC sequences encoding human pancreatic tumour proteins are useful for
 CC stimulating an immune response in a patient and treating pancreatic
 CC cancer in a patient. A host cell that expresses these polynucleotides
 CC is useful for determining the presence of cancer in a patient. A
 CC composition comprising the polynucleotide, its encoded protein, or an
 CC antibody that binds to the protein may be used in the diagnosis,
 CC prevention and/or treatment of diseases, particularly pancreatic
 CC cancer. The sequences of the invention are also useful in pharmaceutical
 CC compositions, e.g. vaccines, for the diagnosis and treatment of
 CC pancreatic cancer. Such compositions may be useful for inhibiting the
 CC development of cancer in a patient, or as markers for the progression
 CC of cancer. The polynucleotide sequences may also be used as probes
 CC or primers for nucleic acid hybridisation assays. ABK44061-ABK44209

CC represent cDNA sequences encoding for human pancreatic tumour proteins.
XX Sequence 561 BP; 121 A; 171 C; 141 G; 126 T; 2 other;
SQ

Query Match 31.9%; Score 33.2; DB 24; Length 561;
Best Local Similarity 61.6%; Pred. No. 0.42; Mismatches 0; Gaps 0;
Matches 53; Conservative 0; Indels 33; Indels 0; Gaps 0;

QY 2 CTGGACCTCCAGCACACCCAGGCCCGCCAGAGAAAGTGGGGCTCTCTGGTGACACAGGTT 61
||||| ||||| | || | ||||| | || | ||||| | ||||| | ||||| | |||||
Db 404 CTGGATCTCCAGATACCAAGGACCCCTGGTGAACCTGGGCAAGCTGGTCTTCAGGCC 463

QY 62 TACCACAATATACAGAGAAATAAGT 87
||| | | | | | |||||
Db 464 CTCGAGGACCTCTGGTGCTATAGGT 489

RESULT 6
ABK44111
ID ABK44111 standard; cDNA; 617 BP.
XX
AC ABK44111;
XX
DT 21-MAY-2002 (first entry)
XX
DE cDNA #51 encoding human pancreatic tumour protein.
XX
KW Human; pancreatic tumour protein; immune response; pancreatic cancer;
KW development of cancer; cancer progression; cytostatic; gene; ss.
XX
OS Homo sapiens.
XX
PN WO200212331-A2.
XX
PD 14-FEB-2002.
XX
PF 06-AUG-2001; 2001WO-US24619.
XX
PR 07-AUG-2000; 2000US-223130P.
PR 30-JAN-2001; 2001US-265447P.
PR 15-MAY-2001; 2001US-291201P.
XX
PA (CORI-) CORIXA CORP.
XX
PI Pyle RA, Xu J, Kalos MD;
XX
DR WPI; 2002-241741/29.
XX
PT Novel polynucleotide encoding pancreatic tumour polypeptides, useful in
PT pharmaceutical compositions, e.g. vaccines, for treating pancreatic
PT cancers
XX
PS Claim 1; Page 127; 167pp; English.
XX
CC The present invention relates to the isolation of cDNA sequences
CC encoding human pancreatic tumour proteins. The polynucleotide
CC sequences encoding human pancreatic tumour proteins are useful for
CC stimulating an immune response in a patient and treating pancreatic
CC cancer in a patient. A host cell that expresses these polynucleotides
CC is useful for determining the presence of cancer in a patient. A
CC composition comprising the polynucleotide, its encoded protein, or an
CC antibody that binds to the protein may be used in the diagnosis,
CC prevention and/or treatment of diseases, particularly pancreatic
CC cancer. The sequences of the invention are also useful in pharmaceutical
CC compositions, e.g. vaccines, for the diagnosis and treatment of
CC pancreatic cancer. Such compositions may be useful for inhibiting the
CC development of cancer in a patient, or as markers for the progression
CC of cancer. The polynucleotide sequences may also be used as probes
CC or primers for nucleic acid hybridisation assays. ABK44061-ABK44209
CC represent cDNA sequences encoding for human pancreatic tumour proteins.
XX
XX Sequence 617 BP; 134 A; 186 C; 157 G; 138 T; 2 other;

Query Match 31.9%; Score 33.2; DB 24; Length 617;
Best Local Similarity 61.6%; Pred. No. 0.43; Mismatches 0; Gaps 0;
Matches 53; Conservative 0; Indels 33; Indels 0; Gaps 0;

QY 2 CTGGACCTCCAGCACACCCAGGCCCGCCAGAGAAAGTGGGGCTCTCTGGTGACACAGGTT 61
||||| ||||| | || | ||||| | || | ||||| | ||||| | ||||| | |||||
Db 404 CTGGATCTCCAGATACCAAGGACCCCTGGTGAACCTGGGCAAGCTGGTCTTCAGGCC 463

QY 62 TACCACAATATACAGAGAAATAAGT 87
||| | | | | | |||||
Db 464 CTCGAGGACCTCTGGTGCTATAGGT 489

RESULT 7
ABK44126
ID ABK44126 standard; cDNA; 648 BP.
XX
AC ABK44126;
XX
DT 21-MAY-2002 (first entry)
XX
DE cDNA #66 encoding human pancreatic tumour protein.
XX
KW Human; pancreatic tumour protein; immune response; pancreatic cancer;
KW development of cancer; cancer progression; cytostatic; gene; ss.
XX
OS Homo sapiens.
XX
PN WO200212331-A2.
XX
PD 14-FEB-2002.
XX
PF 06-AUG-2001; 2001WO-US24619.
XX
PR 07-AUG-2000; 2000US-223130P.
PR 30-JAN-2001; 2001US-265447P.
PR 15-MAY-2001; 2001US-291201P.
XX
PA (CORI-) CORIXA CORP.
XX
PI Pyle RA, Xu J, Kalos MD;
XX
DR WPI; 2002-241741/29.
XX
PT Novel polynucleotide encoding pancreatic tumour polypeptides, useful in
PT pharmaceutical compositions, e.g. vaccines, for treating pancreatic
PT cancers
XX
PS Claim 1; Page 132; 167pp; English.
XX
CC The present invention relates to the isolation of cDNA sequences
CC encoding human pancreatic tumour proteins. The polynucleotide
CC sequences encoding human pancreatic tumour proteins are useful for
CC stimulating an immune response in a patient and treating pancreatic
CC cancer in a patient. A host cell that expresses these polynucleotides
CC is useful for determining the presence of cancer in a patient. A
CC composition comprising the polynucleotide, its encoded protein, or an
CC antibody that binds to the protein may be used in the diagnosis,
CC prevention and/or treatment of diseases, particularly pancreatic
CC cancer. The sequences of the invention are also useful in pharmaceutical
CC compositions, e.g. vaccines, for the diagnosis and treatment of
CC pancreatic cancer. Such compositions may be useful for inhibiting the
CC development of cancer in a patient, or as markers for the progression
CC of cancer. The polynucleotide sequences may also be used as probes
CC or primers for nucleic acid hybridisation assays. ABK44061-ABK44209
CC represent cDNA sequences encoding for human pancreatic tumour proteins.
XX
XX Sequence 648 BP; 147 A; 188 C; 169 G; 142 T; 2 other;

Query Match 31.9%; Score 33.2; DB 24; Length 648;
Best Local Similarity 61.6%; Pred. No. 0.44; Mismatches 0; Gaps 0;
Matches 53; Conservative 0; Indels 33; Indels 0; Gaps 0;

Qy 62 TACCACAATATACAGGAGAAATAAGT 87
 Dd 464 CTCACAGGACCTCCTGGTGCTATAGGT 489

RESULT 9
ABK44128
ID ABK.

DT 21-MAY-2002 (first entry)

DE CDNA #68 encoding human pancreatic tumour protein.

Human; pancreatic tumour protein; immune response; pancreatic cancer development of cancer; cancer progression; cytostatic; gene; ss.

OS Homo sapiens.

PN WO200212331-A2

PD 14-FEB-2002.

06-AUG-2001; 2001WO-US24619.

PR 07-AUG-2000: 2000US-223130P.

PK 30-JAN-2001; 2001032003447E
PB 15-MAY-2001; 200105291201P

PA (CORT-) CORTXA CORP

XX
PT
py]e PA
Ym J
Kajos MD.XX
DP WPT: 2003-2A17A1/20

2A	Novel polynucleotide encoding pancreatic tumour polypeptides, useful in
PT	pharmaceutical compositions, e.g. vaccines, for treating pancreatic
PT	cancers -

PS Claim 1; Page 132; 167pp; English.

The present invention relates to the isolation of cDNA sequences encoding human pancreatic tumour proteins. The polynucleotide sequences encoding human pancreatic tumour proteins are useful for stimulating an immune response in a patient and treating pancreatic cancer in a patient. A host cell that expresses these polynucleotides is useful for determining the presence of cancer in a patient. A composition comprising the polynucleotide, its encoded protein, or an antibody that binds to the protein may be used in the diagnosis, prevention and/or treatment of diseases, particularly pancreatic cancer. The sequences of the invention are also useful in pharmaceutical compositions, e.g. vaccines, for the diagnosis and treatment of pancreatic cancer. Such compositions may be useful for inhibiting the development of cancer in a patient, or as markers for the progression of cancer. The polynucleotide sequences may also be used as probes or primers for nucleic acid hybridisation assays. ABK44061-ABK44209 represent cDNA sequences encoding for human pancreatic tumour proteins.

Sequence 655 BP: 150 A; 189 C; 171 G; 144 T; 1 other;

Query Match 31.9%; Score 33.2; DB 24; Length 655;
Best Local Similarity 61.6%; Pred. No. 0.44;
Matches 53; Conservative 0; Mismatches 33; Indels 0

Qy	2	CTGGACCTCAGCACACCCAGGCCCCAGAGAAGTGGGCGCTCCTGGTGCAACCAGATT	61
Db	404	CTGGATCTCAGGATACCAAGGACCCCCTGGTGAACCTGGGCAAGCTGGTCCTTCAGGCC	463

```
RESULT 10
ABK44129
ID ABK44129 standard; cDNA; 656 BP.
XX
AC ABK44129;
XX
XX 21-MAY-2002 (first entry)
XX
DE DE
KW Human; pancreatic tumour protein; immune response; pancreatic cancer;
KW development of cancer; cancer progression; cytostatic; gene; ss.
XX
OS Homo sapiens.
XX
PN WO200212331-A2.
XX
PD 14-FEB-2002.
XX
PF 06-AUG-2001; 2001WO-US24619.
XX
PR 07-AUG-2000; 2000US-223130P.
PR 30-JAN-2001; 2001US-265447P.
PR 15-MAY-2001; 2001US-291201P.
XX
PA (CORI-) CORIXA CORP.
XX
PI Pyle RA, Xu J, Kalos MD;
XX
DR WPI; 2002-241741/29.
XX
PT Novel polynucleotide encoding pancreatic tumour polypeptides, useful in
PT pharmaceutical compositions, e.g. vaccines, for treating pancreatic
PT cancers -
XX
PS Claim 1; Page 133; 167pp; English.
XX
CC The present invention relates to the isolation of cDNA sequences
CC encoding human pancreatic tumour proteins. The polynucleotide
CC sequences encoding human pancreatic tumour proteins are useful for
CC stimulating an immune response in a patient and treating pancreatic
CC cancer in a patient. A host cell that expresses these polynucleotides
CC is useful for determining the presence of cancer in a patient. A
CC composition comprising the polynucleotide, its encoded protein, or an
CC antibody that binds to the protein may be used in the diagnosis,
CC prevention and/or treatment of diseases, particularly pancreatic
CC cancer. The sequences of the invention are also useful in pharmaceutical
CC compositions, e.g. vaccines, for the diagnosis and treatment of
CC pancreatic cancer. Such compositions may be useful for inhibiting the
CC development of cancer in a patient, or as markers for the progression
CC of cancer. The polynucleotide sequences may also be used as probes
CC or primers for nucleic acid hybridisation assays. ABK44061-ABK44209
CC represent cDNA sequences encoding for human pancreatic tumour proteins.
XX
SQ Sequence 656 BP; 152 A; 188 C; 168 G; 145 T; 3 other;

Query Match 31.9%; Score 33.2; DB 24; Length 656;
Best Local Similarity 61.6%; Pred. No. 0.44;
Matches 53; Conservative 0; Mismatches 33; Indels 0; Gaps 0;

QY 2 CTGGACCTCCAGCACACCCAGGCCGCCAGAGAGTGGGGCTCTCTGTGCACCAAGGTT 61
DB 404 CTGGATCTCCAGGATACCAAGACCCCTGTGTGAACCTGGGCAAGCTGCTCTCAGGCC 463

QY 62 TACCACAATATACAGGAGAAATAAGT 87
DB 464 CTCAGGACCTCTCTGTGCTATAGT 489

RESULT 11
ABK44131
ID ABK44131 standard; cDNA; 658 BP.
XX
AC ABK44131;
XX
XX 21-MAY-2002 (first entry)
XX
DE DE
KW Human; pancreatic tumour protein; immune response; pancreatic cancer;
KW development of cancer; cancer progression; cytostatic; gene; ss.
XX
OS Homo sapiens.
XX
PN WO200212331-A2.
XX
PD 14-FEB-2002.
XX
PF 06-AUG-2001; 2001WO-US24619.
XX
PR 07-AUG-2000; 2000US-223130P.
PR 30-JAN-2001; 2001US-265447P.
PR 15-MAY-2001; 2001US-291201P.
XX
PA (CORI-) CORIXA CORP.
XX
PI Pyle RA, Xu J, Kalos MD;
XX
DR WPI; 2002-241741/29.
XX
PT Novel polynucleotide encoding pancreatic tumour polypeptides, useful in
PT pharmaceutical compositions, e.g. vaccines, for treating pancreatic
PT cancers -
XX
PS Claim 1; Page 133; 167pp; English.
XX
CC The present invention relates to the isolation of cDNA sequences
CC encoding human pancreatic tumour proteins. The polynucleotide
CC sequences encoding human pancreatic tumour proteins are useful for
CC stimulating an immune response in a patient and treating pancreatic
CC cancer in a patient. A host cell that expresses these polynucleotides
CC is useful for determining the presence of cancer in a patient. A
CC composition comprising the polynucleotide, its encoded protein, or an
CC antibody that binds to the protein may be used in the diagnosis,
CC prevention and/or treatment of diseases, particularly pancreatic
CC cancer. The sequences of the invention are also useful in pharmaceutical
CC compositions, e.g. vaccines, for the diagnosis and treatment of
CC pancreatic cancer. Such compositions may be useful for inhibiting the
CC development of cancer in a patient, or as markers for the progression
CC of cancer. The polynucleotide sequences may also be used as probes
CC or primers for nucleic acid hybridisation assays. ABK44061-ABK44209
CC represent cDNA sequences encoding for human pancreatic tumour proteins.
XX
SQ Sequence 658 BP; 151 A; 190 C; 172 G; 144 T; 1 other;
```

```
XX
AC ABK44131;
XX
DT 21-MAY-2002 (first entry)
XX
DE DE
KW Human; pancreatic tumour protein; immune response; pancreatic cancer;
KW development of cancer; cancer progression; cytostatic; gene; ss.
XX
OS Homo sapiens.
XX
PN WO200212331-A2.
XX
PD 14-FEB-2002.
XX
PF 06-AUG-2001; 2001WO-US24619.
XX
PR 07-AUG-2000; 2000US-223130P.
PR 30-JAN-2001; 2001US-265447P.
PR 15-MAY-2001; 2001US-291201P.
XX
PA (CORI-) CORIXA CORP.
XX
PI Pyle RA, Xu J, Kalos MD;
XX
DR WPI; 2002-241741/29.
XX
PT Novel polynucleotide encoding pancreatic tumour polypeptides, useful in
PT pharmaceutical compositions, e.g. vaccines, for treating pancreatic
PT cancers -
XX
PS Claim 1; Page 133; 167pp; English.
XX
CC The present invention relates to the isolation of cDNA sequences
CC encoding human pancreatic tumour proteins. The polynucleotide
CC sequences encoding human pancreatic tumour proteins are useful for
CC stimulating an immune response in a patient and treating pancreatic
CC cancer in a patient. A host cell that expresses these polynucleotides
CC is useful for determining the presence of cancer in a patient. A
CC composition comprising the polynucleotide, its encoded protein, or an
CC antibody that binds to the protein may be used in the diagnosis,
CC prevention and/or treatment of diseases, particularly pancreatic
CC cancer. The sequences of the invention are also useful in pharmaceutical
CC compositions, e.g. vaccines, for the diagnosis and treatment of
CC pancreatic cancer. Such compositions may be useful for inhibiting the
CC development of cancer in a patient, or as markers for the progression
CC of cancer. The polynucleotide sequences may also be used as probes
CC or primers for nucleic acid hybridisation assays. ABK44061-ABK44209
CC represent cDNA sequences encoding for human pancreatic tumour proteins.
XX
SQ Sequence 658 BP; 151 A; 190 C; 172 G; 144 T; 1 other;

Query Match 31.9%; Score 33.2; DB 24; Length 658;
Best Local Similarity 61.6%; Pred. No. 0.44;
Matches 53; Conservative 0; Mismatches 33; Indels 0; Gaps 0;

QY 2 CTGGACCTCCAGCACACCCAGGCCGCCAGAGAGTGGGGCTCTCTGTGCACCAAGGTT 61
DB 404 CTGGATCTCCAGGATACCAAGACCCCTGTGTGAACCTGGGCAAGCTGCTCTCAGGCC 463

QY 62 TACCACAATATACAGGAGAAATAAGT 87
DB 464 CTCAGGACCTCTCTGTGCTATAGT 489

RESULT 12
ABK44112
ID ABK44112 standard; cDNA; 659 BP.
XX
AC ABK44112;
XX
XX 21-MAY-2002 (first entry)
```

XX cDNA #52 encoding human pancreatic tumour protein.
 XX Human; pancreatic tumour protein; immune response; pancreatic cancer;
 KW development of cancer; cancer progression; cytostatic; gene; ss.
 KW Homo sapiens.
 XX WO200212331-A2.
 XX 14-FEB-2002.
 XX 06-AUG-2001; 2001WO-US24619.
 XX 07-AUG-2000; 2000US-223130P.
 PR 30-JAN-2001; 2001US-265447P.
 PR 15-MAY-2001; 2001US-291201P.
 XX (CORI-) CORIXA CORP.
 XX Pyle RA, Xu J, Kalos MD;
 XX WPI; 2002-241741/29.
 XX Novel polynucleotide encoding pancreatic tumour polypeptides, useful in
 PT pharmaceutical compositions, e.g. vaccines, for treating pancreatic
 PT cancers -
 XX Claim 1; Page 127-128; 167pp; English.
 XX The present invention relates to the isolation of cDNA sequences
 CC encoding human pancreatic tumour proteins. The polynucleotide
 CC sequences encoding human pancreatic tumour proteins are useful for
 CC stimulating an immune response in a patient and treating pancreatic
 CC cancer in a patient. A host cell that expresses these polynucleotides
 CC is useful for determining the presence of cancer in a patient. A
 CC antibody that binds to the protein may be used in the diagnosis,
 CC prevention and/or treatment of diseases, particularly pancreatic
 CC cancer. The sequences of the invention are also useful in pharmaceutical
 CC compositions, e.g. vaccines, for the diagnosis and treatment of
 CC pancreatic cancer. Such compositions may be useful for inhibiting the
 CC development of cancer in a patient, or as markers for the progression
 CC of cancer. The polynucleotide sequences may also be used as probes
 CC or primers for nucleic acid hybridisation assays. ABK44061-ABK44209
 CC represent cDNA sequences encoding for human pancreatic tumour proteins.
 XX Sequence 659 BP; 152 A; 189 C; 174 G; 143 T; 1 other;
 SQ
 Query Match 31.9%; Score 33.2; DB 24; Length 659;
 Best Local Similarity 61.6%; Pred. No. 0.44;
 Matches 53; Conservative 0; Mismatches 33; Indels 0; Gaps 0;
 QY 2 CTGGACCTCCAGCACACCCAGGCCCCAGAGAAGTGGGCGCTCTGGTGCCACCCAGGTT 61
 Db 404 CTGGATCTCCAGGATACCAAGGACCCCTGGTGAACCTGGGCAAGCTGGCTTCAGGCC 463
 QY 62 TACCACAATATACAGGAGAAATAGT 87
 Db 464 CTCAGGACCTCTGGTGCTATAGT 489
 RESULT 13
 ID ABK44065/C
 XX ABK44065 standard; cDNA; 780 BP.
 XX AC ABK44065;
 XX 21-MAY-2002 (first entry)
 XX cDNA #5 encoding human pancreatic tumour protein.
 DE Human; pancreatic tumour protein; immune response; pancreatic cancer;
 KW development of cancer; cancer progression; cytostatic; gene; ss.
 XX Homo sapiens.

KW development of cancer; cancer progression; cytostatic; gene; ss.
 XX Homo sapiens.
 OS WO200212331-A2.
 PN 14-FEB-2002.
 XX 06-AUG-2001; 2001WO-US24619.
 XX 07-AUG-2000; 2000US-223130P.
 PR 30-JAN-2001; 2001US-265447P.
 PR 15-MAY-2001; 2001US-291201P.
 XX (CORI-) CORIXA CORP.
 XX Pyle RA, Xu J, Kalos MD;
 XX WPI; 2002-241741/29.
 XX Novel polynucleotide encoding pancreatic tumour polypeptides, useful in
 PT pharmaceutical compositions, e.g. vaccines, for treating pancreatic
 PT cancers -
 XX Claim 1; Page 111; 167pp; English.
 XX The present invention relates to the isolation of cDNA sequences
 CC encoding human pancreatic tumour proteins. The polynucleotide
 CC sequences encoding human pancreatic tumour proteins are useful for
 CC stimulating an immune response in a patient and treating pancreatic
 CC cancer in a patient. A host cell that expresses these polynucleotides
 CC is useful for determining the presence of cancer in a patient. A
 CC antibody that binds to the protein may be used in the diagnosis,
 CC prevention and/or treatment of diseases, particularly pancreatic
 CC cancer. The sequences of the invention are also useful in pharmaceutical
 CC compositions, e.g. vaccines, for the diagnosis and treatment of
 CC pancreatic cancer. Such compositions may be useful for inhibiting the
 CC development of cancer in a patient, or as markers for the progression
 CC of cancer. The polynucleotide sequences may also be used as probes
 CC or primers for nucleic acid hybridisation assays. ABK44061-ABK44209
 CC represent cDNA sequences encoding for human pancreatic tumour proteins.
 XX Sequence 780 BP; 163 A; 211 C; 218 G; 171 T; 17 other;
 SQ
 Query Match 31.9%; Score 33.2; DB 24; Length 780;
 Best Local Similarity 61.6%; Pred. No. 0.47;
 Matches 53; Conservative 0; Mismatches 33; Indels 0; Gaps 0;
 QY 2 CTGGACCTCCAGCACACCCAGGCCCCAGAGAAGTGGGCGCTCTGGTGCCACCCAGGTT 61
 Db 392 CTGGATCTCCAGGATACCAAGGACCCCTGGTGAACCTGGGCAAGCTGGCTTCAGGCC 333
 QY 62 TACCACAATATACAGGAGAAATAGT 87
 Db 332 CTCAGGACCTCTCTGGTGCTATAGT 307
 RESULT 14
 ID ABK44088
 XX ABK44088 standard; cDNA; 823 BP.
 XX AC ABK44088;
 XX 21-MAY-2002 (first entry)
 XX cDNA #28 encoding human pancreatic tumour protein.
 DE Human; pancreatic tumour protein; immune response; pancreatic cancer;
 KW development of cancer; cancer progression; cytostatic; gene; ss.
 XX Homo sapiens.

```
PN WO200212331-A2.
XX
PD 14-FEB-2002.
XX
PF 06-AUG-2001; 2001WO-US24619.
XX
PR 07-AUG-2000; 2000US-223130P.
PR 30-JAN-2001; 2001US-265447P.
PR 15-MAY-2001; 2001US-291201P.
XX
PA (CORI-) CORIXA CORP.
XX
PI Pyle RA, Xu J, Kalos MD;
XX
DR WPI; 2002-241741/29.
XX
PT Novel polynucleotide encoding pancreatic tumour polypeptides, useful in
PT pharmaceutical compositions, e.g. vaccines, for treating pancreatic
PT cancers.
XX
PS Claim 1; Page 119; 167pp; English.
XX
CC The present invention relates to the isolation of cDNA sequences
CC encoding human pancreatic tumour proteins. The polynucleotide
CC sequences encoding human pancreatic tumour proteins are useful for
CC stimulating an immune response in a patient and treating pancreatic
CC cancer in a patient. A host cell that expresses these polynucleotides
CC is useful for determining the presence of cancer in a patient. A
CC composition comprising the polynucleotide, its encoded protein, or an
CC antibody that binds to the protein may be used in the diagnosis,
CC prevention and/or treatment of diseases, particularly pancreatic
CC cancer. The sequences of the invention are also useful in pharmaceutical
CC compositions, e.g. vaccines, for the diagnosis and treatment of
CC pancreatic cancer. Such compositions may be useful for inhibiting the
CC development of cancer in a patient, or as markers for the progression
CC of cancer. The polynucleotide sequences may also be used as probes
CC or primers for nucleic acid hybridisation assays. ABK44061-ABK44209
CC represent cDNA sequences encoding for human pancreatic tumour proteins.
XX
SQ Sequence 823 Bp; 182 A; 228 C; 237 G; 173 T; 3 other;

Query Match 31.9%; Score 33.2; DB 24; Length 823;
Best Local Similarity 61.6%; Pred. No. 0.48;
Matches 53; Conservative 0; Mismatches 33; Indels 0; Gaps 0;

QY 2 CTGGACCTCCAGCACCCAGGCCCCAGAGAGAGTGCGGCGCTCTCTGTGCACGAGTT 61
Db 413 CTGGATCTCCAGGATACCAAGGACCCCTGTGTAACCTGGGCAAGCTGTCTTCAAGGCC 472

QY 62 TACCACAATATACAGGAGAAATAAGT 87
Db 473 CTCAGGACCTCTGTGTCTATAGT 498

RESULT 15
AAF44902
ID AAF44902 standard; cDNA; 994 Bp.
XX
AC AAF44902;
XX
DT 28-MAR-2001 (first entry)
XX
DE Human breast cancer related protein coding sequence SEQ ID NO: 58.
XX
KW Human; breast cancer; diagnosis; therapy; vaccine; ss.
XX
OS Homo sapiens.
XX
PN WO200078960-A2.
XX
PD 28-DEC-2000.
XX
PF 23-JUN-2000; 2000WO-US17536.
XX
PT
```

```
XX 23-JUN-1999; 99US-0140903.
PR 12-OCT-1999; 99US-0158980.
XX
PA (CORI-) CORIXA CORP.
XX
PI Yuqiu J, Mitcham JL;
XX
DR WPI; 2001-041426/05.
XX
PT New polynucleotides encoding breast tumor specific proteins, useful for
PT prevention, treatment and diagnosis of breast cancer.
XX
PS Claim 25; Page 138; 165pp; English.
XX
CC The present invention provides the coding sequences for a number of
CC breast cancer related proteins. These can be used in vaccinations
CC against, diagnosis of and treatment of cancer, particularly breast
CC cancer.
XX
SQ Sequence 994 Bp; 234 A; 260 C; 262 G; 237 T; 1 other;

Query Match 31.9%; Score 33.2; DB 22; Length 994;
Best Local Similarity 61.6%; Pred. No. 0.51;
Matches 53; Conservative 0; Mismatches 33; Indels 0; Gaps 0;

QY 2 CTGGACCTCCAGCACCCAGGCCCCAGAGAGAGTGCGGCGCTCTCTGTGCACGAGTT 61
Db 685 CTGGATCTCCAGGATACCAAGGACCCCTGTGTAACCTGGGCAAGCTGTCTTCAAGGCC 744

QY 62 TACCACAATATACAGGAGAAATAAGT 87
Db 745 CTCAGGACCTCTGTGTCTATAGT 770

Search completed: February 19, 2003, 22:56:18
Job time : 40.9937 secs
```

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 19, 2003, 23:16:26 ; Search time 331.581 Seconds
(without alignments)
9128.055 Million cell updates/sec

Title: US-09-997-610-1-copy_53_156

Perfect score: 104

Sequence: 1 gctgacctccagcacccc.....agtgaatgacaaaatgcc 104

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.*

1: gb_ba.*

2: gb_htg.*

3: gb_in.*

4: gb_om.*

5: gb_ov.*

6: gb_pat.*

7: gb_ph.*

8: gb_pl.*

9: gb_pr.*

10: gb_ro.*

11: gb_sts.*

12: gb_sy.*

13: gb_un.*

14: gb_vi.*

15: em_ba.*

16: em_fun.*

17: em_hum.*

18: em_in.*

19: em_mu.*

20: em_om.*

21: em_or.*

22: em_ov.*

23: em_pat.*

24: em_ph.*

25: em_pi.*

26: em_ro.*

27: em_sts.*

28: em_un.*

29: em_vi.*

30: em_htg_hum.*

31: em_htg_inv.*

32: em_htg_other.*

33: em_htg_mus.*

34: em_htg_pln.*

35: em_htg_rod.*

36: em_htg_mam.*

37: em_htg_vrt.*

38: em_sy.*

39: em_htgo_hum.*

40: em_htgo_mus.*

41: em_htgo_other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	60	57.7	145880	9	HS302D9
2	40.6	39.0	5851	3	AF282902
3	37.4	36.0	287	3	AF282902 Hydra vul
4	36.8	35.4	3612	10	OOS420010
5	36.8	35.4	6512	6	MMCOLA4
6	36.8	35.4	6512	6	AX306181
7	36.2	34.8	694	3	MUSCOLIA4A
8	36.2	34.8	36532	3	HSNCOL1
9	36.2	34.8	269619	3	CEF5781
10	35.6	34.2	11942	1	CEY51H4A
11	35.2	33.8	1993	3	AE010554
12	34.8	33.5	3734	14	BMCOLGMR
13	34.8	33.5	6522	3	U58736
14	34.2	32.9	31110	3	CBRG1AK24
15	34	32.7	3234	9	HSC3AIR
16	34	32.7	87849	3	AC084453
17	33.8	32.5	1776	10	RNU57362
18	33.4	32.1	97083	9	AC010289
19	33.4	32.1	107655	2	AC084840
20	33.4	32.1	147416	9	AC005549
21	33.4	32.1	156000	2	AC060815
22	33.4	32.1	159694	2	AC098586
23	33.4	32.1	162214	9	AC093809
24	33.4	32.1	177941	9	AL445423
25	33.4	32.1	188460	9	AC010542
26	33.4	32.1	198876	2	AC084810
27	33.4	32.1	316776	2	AL603862
28	33.2	31.9	994	6	AX067354
29	33.2	31.9	1881	6	A60690
30	33.2	31.9	1881	6	AR123971
31	33.2	31.9	3902	9	BC028178
32	33.2	31.9	4428	6	AX146422
33	33.2	31.9	4428	6	AX146424
34	33.2	31.9	4428	6	AX146430
35	33.2	31.9	5460	6	AX302553
36	33.2	31.9	5460	6	AX393295
37	33.2	31.9	5460	6	E10600
38	33.2	31.9	5460	9	HSCOL3AI
39	33.2	31.9	68808	2	AC022708
40	33.2	31.9	96799	9	AC104647
41	33.2	31.9	105936	3	AC084440
42	33.2	31.9	198491	2	AC022249
43	33	31.7	2247	3	CBUR4501
44	33	31.7	4235	10	BC016479
45	33	31.7	9448	3	CEU22327

ALIGNMENTS

RESULT 1	HS302D9	145880 bp	DNA	linear	PRI 12-DEC-1999
LOCUS	Human DNA sequence from clone RPI-302D9 on chromosome 22				Contains
DEFINITION	GS55, complete sequence.				
ACCESSION	282198				
VERSION	282198.2	GI:6572207			
KEYWORDS	HTG.				
SOURCE	Homo sapiens.				
ORGANISM	Homo sapiens.				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
TITLE	1 (bases 1 to 145880)				
	Bridgeman,A.				
	Direct Submission				

JOURNAL

Submitted (08-DEC-1999) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: hamquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk

COMMENT

On Dec 13, 1999 this sequence version replaced gi:3164067. During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

This sequence has been finished according to sequence map criteria as follows. An attempt is made to resolve all sequencing problems, such as compressions and repeats, but not necessarily within known annotated human repeat sequence elements (e.g. Alu). Where the sequence is ambiguous, there is an annotation using the 'unsure' feature key.

The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em, EMBL; Sw, SWISSPROT; Tr, TREMBL; Wp, WORMPEP; Information on the WORMPEP database can be found at

http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence was generated from part of bacterial clone contigs of human chromosome 22, constructed by the Sanger Centre Chromosome 22 Mapping Group. Further information can be found at

<http://www.sanger.ac.uk/HGP/Chr22>
Park Cancer Institute by the group of Pieter de Jong. For further details see <http://bacpac.med.buffalo.edu/> VECTOR: pCYPAC2

This sequence is the entire insert of clone RPI-302D9 The true left end of clone CTA-282F2 is at 69682 in this sequence. The true right end of clone CTA-415G2 is at 55167 in this sequence.

FEATURES

```

source
    1..145880
        /organism="Homo sapiens"
        /db_xref="taxon:9606"
        /chromosome="22"
        /clone="RPI-302D9"
        /clone_1lib="RPI-1"
    188..245
        /note="MER3 repeat: matches 144..209 of consensus"
    246..571
        /note="AluX repeat: matches 1..312 of consensus"
    572..759
        /note="MER3 repeat: matches 1..144 of consensus"
    783..933
        /note="MERSA repeat: matches 26..187 of consensus"
    1033..1336
        /note="AluSp repeat: matches 1..299 of consensus"
    1450..1583
        /note="MIR repeat: matches 24..160 of consensus"
    1687..1752
        /note="L2 repeat: matches 2593..2661 of consensus"
    2350..2660
        /note="AluSc repeat: matches 3..309 of consensus"
    2684..2981
        /note="AluSp repeat: matches 2..300 of consensus"
    3323..3343
        /note="MLTIE repeat: matches 116..136 of consensus"
    3344..3652
        /note="AluY repeat: matches 1..309 of consensus"
    3653..3928
        /note="MLTIE repeat: matches 136..359 of consensus"
    3929..4278
        /note="THEIB repeat: matches 3..364 of consensus"
    4279..4485
        /note="MLTIE repeat: matches 359..568 of consensus"
    5073..5176
        /note="52 copies 2 mer ct 78 conserved"
    5181..5491
        /note="AluNb repeat: matches 1..311 of consensus"
    6369..6485
        /note="L2 repeat: matches 2579..2705 of consensus"
    6647..6685
        /note="MADE1 repeat: matches 1..23 of consensus"
    6686..6987
        /note="AluX repeat: matches 1..302 of consensus"
    6988..7036
        /note="MADE1 repeat: matches 23..77 of consensus"
    7482..7754
        /note="AluJb repeat: matches 9..290 of consensus"
    7775..8060
        /note="AluJo repeat: matches 1..295 of consensus"
    8414..8551
        /note="L2 repeat: matches 2553..2706 of consensus"
    8914..9030
        /note="MIR repeat: matches 147..262 of consensus"
    9110..9280
        /note="MIR repeat: matches 91..262 of consensus"
    9283..9412
        /note="MIR repeat: matches 15..144 of consensus"
    9521..9679
        /note="FAM repeat: matches 3..161 of consensus"
    9820..10225
        /note="MSTB repeat: matches 2..425 of consensus"
    complement(10179..10678)
        /note="match: GSS: Em:B56592"
    complement(10204..10728)
        /note="match: GSS: Em:A0701486"
    complement(10249..10706)
        /note="match: GSS: Em:AQ225495"
    10312..10383
        /note="MIR repeat: matches 79..150 of consensus"
    10718..11310
        /note="match: GSS: Em:B14024"
    10784..11201
        /note="match: GSS: Em:B43656"
    11838..11946
        /note="MIR repeat: matches 20..137 of consensus"
    12174..12445
        /note="L2 repeat: matches 1988..2275 of consensus"
    12444..12642
        /note="MIR repeat: matches 63..241 of consensus"
    13017..13369
        /note="match: STS: Em:G49301"
    13331..13397
        /note="MIR repeat: matches 174..244 of consensus"
    13398..13698
        /note="AluSp repeat: matches 1..302 of consensus"
    13699..13810
        /note="MIR repeat: matches 76..174 of consensus"
    13806..13919
        /note="MIR repeat: matches 77..189 of consensus"
    13945..14060
        /note="MIR repeat: matches 24..142 of consensus"
    14061..14367
        /note="AluY repeat: matches 1..301 of consensus"
    14368..14452
        /note="MIR repeat: matches 141..225 of consensus"
    14589..14679
        /note="MIR repeat: matches 173..262 of consensus"
    14597..15201
        /note="match: GSS: Em:AQ553482"
    14616..15060
        /note="match: GSS: Em:AQ370601"
    14868..15040
        /note="MIR repeat: matches 49..233 of consensus"
    15071..15188
        /note="L2 repeat: matches 2112..2239 of consensus"
    15304..15399
        /note="MLTIB repeat: matches 1..99 of consensus"
    15490..15662
        /note="AluSg1 repeat: matches 2..114 of consensus"
    15669..15727
        /note="MLTIB repeat: matches 119..178 of consensus"
    15728..16027

```

```
repeat_region /note="AluSc repeat: matches 1. .299 of consensus"
16028. .16245
/note="MT1B repeat: matches 178. .390 of consensus"
repeat_region 16546. .16854
/note="AluY repeat: matches 1. .300 of consensus"
repeat_region 18296. .18323
/note="MSTR repeat: matches 2. .29 of consensus"
repeat_region 18324. .18392
/note="MER66-internal repeat: matches 4919. .4993 of
consensus"
repeat_region 18393. .18712
/note="AluJb repeat: matches 1. .311 of consensus"
repeat_region 18713. .19133
/note="MER66-internal repeat: matches 4548. .4919 of
consensus"
misc_feature complement(18872. .19230)
misc_feature 19251. .19719 GSS: Em:AQ005063"
/note="match: GSS: Em:B14179"
repeat_region 19537. .20290
/note="HERVH21 repeat: matches 4657. .5784 of consensus"
repeat_region 20317. .20382
/note="33 copies 2 mer ta 68 conserved"
repeat_region 20513. .20666
/note="77 copies 2 mer tt 70 conserved"
repeat_region 20682. .21008
/note="AluSg1 repeat: matches 1. .306 of consensus"
repeat_region 21239. .21553
/note="HUEP3 repeat: matches 4410. .4713 of consensus"
repeat_region 21882. .22254
/note="THE1B repeat: matches 1. .364 of consensus"
repeat_region 22302. .22537
/note="MER66-internal repeat: matches 2186. .2417 of
consensus"
repeat_region 22538. .22850
/note="AluSp repeat: matches 1. .313 of consensus"
repeat_region 22851. .23801
/note="MER66-internal repeat: matches 1210. .2186 of
consensus"
repeat_region 23905. .23989
/note="MER66-internal repeat: matches 3017. .3102 of
consensus"
Query Match 57.7%; Score 60; DB 9; Length 145880;
Best Local Similarity 100.0%; Pred. No. 4e-08;
Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCTGGACCTCCACACACCCAGCCAGAGAACTGGGCTCTCTGTCACACAGGT 60
|||||
Db 34837 GCTGGACCTCCACACACCCAGCCAGAGAACTGGGCTCTCTGTCACACAGGT 34896

RESULT 2
AF282902 5851 bp mRNA linear INV 17-DEC-2000
LOCUS Hydra vulgaris type IV collagen alpha 1 chain precursor, mRNA,
DEFINITION complete cds.
ACCESSION AF282902
VERSION AF282902.1 GI:11875611
KEYWORDS Hydra vulgaris.
SOURCE Hydra vulgaris.
ORGANISM Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydroids; Anthomedusae;
Hydridae; Hydra.
REFERENCE 1 (bases 1 to 5851)
AUTHORS Fowler,S.J., Jose,S., Zhang,X., Deutzmann,R., Sarraz,M.P. Jr. and
Boot-Handford,R.P.
TITLE Characterization of hydra type IV collagen. Type IV collagen is
essential for head regeneration and its expression is up-regulated
upon exposure to glucose
J. Biol. Chem. 275 (50), 39589-39599 (2000)
JOURNAL 20564332
MEDLINE PUBMED 10956657
REFERENCE 2 (bases 1 to 5851)
```

```
AUTHORS Fowler,S.J. and Boot-Handford,R.P.
TITLE Direct Submission
JOURNAL Submitted (27-JUN-2000) School of Biological Sciences, University
of Manchester, Oxford Road, Manchester M13 9PT, UK
FEATURES
source Location/Qualifiers
1. .5851
/organism="Hydra vulgaris"
/db_xref="taxon:6087"
157. .5328
/codon_start=1
/product="type IV collagen alpha 1 chain precursor"
/protein_id="AAG40729.1"
/db_xref="GI:11875612"
/translat="MNHTTKWTIWIIVLFTLHSLVFTQLCGGGCNCQSNVCVVGKRG
DRSGVGLPGFKGPIGEPFGPGGKAGKGDGAGVQGLQGEVKGKVPCTPGI
PGSHRPGDGEKPGKPGKNGTGDIGPPGKDGVOGSLGPGIPGPGPGS
TATKLKQKGEKPGKSDGPGAGDPTKDGKESGKAGRMGPGQLRKGKAGDSNII
FGERKGDIGLPGPGDCNCSVTGELINIQGPKGEQKKGKDGKQKQKPGPGQA
GEAGDQKGEKDGKEIGSAGSGIOGEGKDGATGAGTKEGPDGSLGSPQKG
DHGTGSEKGVDRGKQGESITGPGQIGKGQKGEKGEKGEKGEKGEKGEK
PGPLKGLKEIGMTGTPGSGPKTGAEMKGPICGAGORGETGSKSGRPGSGVGS
PGMDGPGOIGPPEPCIPGPGPGNPGVNDLGTVLPPGPDGPGQIQGVGP
QCSPIRGDKDACKSCPSQGEKGVAGDGLPGANGKGDGDEKPGLEKSGVGP
GESLTPGQPIGEGKGTGAKNGKRGSDRIVQGERGKGTDMGPSGSGFKGQ
GVDPGTDGPGQGLQPGDGRGAPKSGIPNNGEKGKGNKGDGLRGFKGERG
DEPGLKAGSGKQKGEKDGATGICAGKNGKGDGDPKGEKGLQGLQGISGP
GEPGLKAGSGKQKGEKDGATGICAGKNGKGDGDPKGEKGLQGLQGISGP
IKGEPGLKAGSGKQKGEKDGATGICAGKNGKGDGDPKGEKGLQGLQGISGP
SVVLPGDPEKSGKQKGEKDGATGICAGKNGKGDGDPKGEKGLQGLQGISGP
EKVGTGTDGDEKGEKGEKGTGKGTGTDGIGLQKGERGERGERGERGERGER
KQVGPAGSGKQKGEKDGATGICAGKNGKGDGDPKGEKGLQGLQGISGP
GSLKGNPGEPGPGTGPAGKDYLDLEVPGLGEKPTKGEKGDGKGDGIG
ETCSQTKGEKGVKSGEKEGLEKGLGKGLGKGLGKGLGKGLGKGLGKGLG
DRKEGAGQKGEKGEKGEKGEKGEKGEKGEKGEKGEKGEKGEKGEKGEK
SIPGEPKDSQKGEKGEKGEKGEKGEKGEKGEKGEKGEKGEKGEKGEKGE
DIGPGETGIDGQKGEKGEKGEKGEKGEKGEKGEKGEKGEKGEKGEKGE
SGSGDFGLVHKSQIKVPSCPMOTWEGYSFLVAGNERAFGLQGLQGISCLP
FSTMPFLFCIDIONKVSNDYSFNLSTAEKPKAEPSSGADLNYISRLVCAPSH
VLAVHQSOLDPKCPDGNLWTFSLMTNKSAGQSGQLSSSGSCLDFEVRNPI
ECHGRGTCTWYPTLSFNLSTIGESNMFOVPKFEILERNLKRVSACVMKSV"
157. .228
sig_peptide
misc_feature 229. .276
/note="Region: non-collagenous domain"
misc_feature 277. .4641
/note="Region: contains collagenous domain"
misc_feature 526. .528
/note="N-glycosylation"
/evidence=not_experimental
misc_feature 4642. .5325
/note="Region: non-collagenous domain"
BASE COUNT 1953 a 942 c 1638 g 1318 t
ORIGIN

Query Match 39.0%; Score 40.6; DB 3; Length 5851;
Best Local Similarity 64.2%; Pred. No. 0.039;
Matches 61; Conservative 0; Mismatches 34; Indels 0; Gaps 0;

QY 4 GGACCTCCAGCACACCCAGCCAGAGAACTGGGCTCTCTGTCACACAGTTTA 63
|||||
Db 2065 GGACCTCAAGGACTCAAGGCTCAAGGTGATCGTGTCTCTGTTAAATCAGGTATA 2124

QY 64 CCACATATACAGCAAAATAGTGAATGACAAA 98
|||||
Db 2125 CCAGGCAATACGAGAGAAAGGTGAAGGAGGAAA 2159

RESULT 3
OOS420010
LOCUS Ostertagia ostertagi partial mRNA for collagen (col gene).
DEFINITION
```

287 bp mRNA linear INV 22-NOV-2001

SOURCE	Hydra sp.	REFERENCE	Hydra sp.	REFERENCE	Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydrozoa; Anthomedusae;
ORGANISM	Hydra sp.	AUTHORS	Hydridae; Hydra	TITLE	Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE	1 (bases 1 to 694)	AUTHORS	none.	TITLE	Genome sequence of the nematode C. elegans: a platform for
AUTHORS	Direct Submission	AUTHORS	none.	TITLE	investigating biology. The C. elegans Sequencing Consortium
JOURNAL	Submitted (22-JUL-1991) E.M. Kurz, University of Munich, Dept of	AUTHORS	none.	TITLE	Science 282 (5396), 2012-2018 (1998)
TITLE	Zoology, Luisenstr 14, Munich 2, GERMANY	AUTHORS	none.	TITLE	99069613
JOURNAL	2 (bases 1 to 694)	AUTHORS	none.	TITLE	9851916
AUTHORS	Kurz.E.M., Holstein.T.W., Petri.B.M., Engel.J. and David.C.N.	AUTHORS	none.	TITLE	The C.elegans Sequencing Consortium.
JOURNAL	Mini-collagens in hydra nematocytes	AUTHORS	none.	TITLE	2 (bases 1 to 36532)
MEDLINE	J. Cell Biol. 115 (4), 1159-1169 (1991)	AUTHORS	none.	TITLE	Sims.M.A.
PUBMED	92064646	AUTHORS	none.	TITLE	Direct Submission
COMMENT	See also X61045-X61048.	AUTHORS	none.	TITLE	Submitted (09-AUG-1996) Nematode Sequencing Project, Sanger
FEATURES	Location/Qualifiers	AUTHORS	none.	TITLE	Institute, Hinxton, Cambridge CB10 1SA, England and Department of
source	1..694	AUTHORS	none.	TITLE	Genetics, Washington University, St. Louis, MO 63110, USA. E-mail:
gene	/organism="Hydra sp."	AUTHORS	none.	TITLE	jes@sanger.ac.uk or rw@nematode.wustl.edu
mRNA	/strain="SF1"	AUTHORS	none.	TITLE	Coding sequences below are predicted from computer analysis, using
CDS	/db_xref="taxon:6086"	AUTHORS	none.	TITLE	predictions from Genefinder (P. Green, U. Washington), and other
gene	/clone="3. 2/55"	AUTHORS	none.	TITLE	available information.
mRNA	/clone_lib="lambda gt10, I-cellspec."	AUTHORS	none.	TITLE	Current sequence finishing criteria for the C. elegans genome
CDS	14..676	AUTHORS	none.	TITLE	sequencing consortium are that all bases are either sequenced
gene	/gene="N-COL 1"	AUTHORS	none.	TITLE	unambiguously on both strands, or on a single strand with both a
mRNA	<14..676	AUTHORS	none.	TITLE	dye primer and dye terminator reaction, from distinct subclones.
CDS	/gene="N-COL 1"	AUTHORS	none.	TITLE	Exceptions are indicated by an explicit note.
gene	/evidence="experimental"	AUTHORS	none.	TITLE	IMPORTANT: This sequence is not the entire insert of clone F57B1.
mRNA	14..463	AUTHORS	none.	TITLE	It may be shorter because we only sequence overlapping sections
CDS	/gene="N-COL 1"	AUTHORS	none.	TITLE	once, or longer because we arrange for a small overlap between
gene	/codon_start=1	AUTHORS	none.	TITLE	neighbouring submissions.
mRNA	/product="mini-collagen"	AUTHORS	none.	TITLE	The true left end of clone F57B1 is at 1 in this sequence. The true
CDS	/protein_id="CAA43379.1"	AUTHORS	none.	TITLE	right end of clone F57B1 is at 5273 in
gene	/db_xref="GI:9447"	AUTHORS	none.	TITLE	sequence Z78060.
mRNA	/db_xref="SPTREMBL:Q00484"	AUTHORS	none.	TITLE	The true left end of clone C34D1 is at 35789 in this sequence. The
CDS	/translation="MAMRLVLACLVGLVGAATQAKTLEMLKRDANPGSGYCPSPVCAPA	AUTHORS	none.	TITLE	start of this sequence (1..104) overlaps with the end of sequence
gene	CAVCCVPP	AUTHORS	none.	TITLE	Z78065.
mRNA	GPFGPGAGGGGLPGAPAPPP	AUTHORS	none.	TITLE	The end of this sequence (35789..36532) overlaps with the start of
CDS	14..73	AUTHORS	none.	TITLE	sequence Z78060.
gene	/gene="N-COL 1"	AUTHORS	none.	TITLE	For a graphical representation of this sequence and its analysis
mRNA	74..460	AUTHORS	none.	TITLE	see: - http://wormbase.sanger.ac.uk/perl/ace/elegans/seq/sequence?
CDS	/product="mini-collagen"	AUTHORS	none.	TITLE	name=F57B1
gene	167..235	AUTHORS	none.	TITLE	IMPORTANT: This sequence is NOT necessarily the entire insert of
mRNA	/note="proline repeat"	AUTHORS	none.	TITLE	the specified clone. It may be shorter because we only sequence
CDS	248..373	AUTHORS	none.	TITLE	overlapping sections once, or longer because we arrange for a small
gene	/gene="N-COL 1"	AUTHORS	none.	TITLE	overlap between neighbouring submissions.
mRNA	/note="collagen-like region"	AUTHORS	none.	TITLE	FEATURES
CDS	218 a 197 c 123 g 156 t	AUTHORS	none.	TITLE	Location/Qualifiers
gene	BASE COUNT	AUTHORS	none.	TITLE	1..36532
mRNA	ORIGIN	AUTHORS	none.	TITLE	/organism="Caenorhabditis elegans"
CDS	Query Match 34.88; Score 36.2; DB 3; Length 694;	AUTHORS	none.	TITLE	/db_xref="taxon:6239"
gene	Best Local Similarity 62.98; Pred. No. 0.74; Indels 0; Gaps 0;	AUTHORS	none.	TITLE	/chromosome="v"
mRNA	Matches 56; Conservative 0; Mismatches 33; Indels 0; Gaps 0;	AUTHORS	none.	TITLE	/clone="F57B1"
CDS	Qy 2 CTGGACCTCCAGCACACCCAGGCCCCAGAGAGAGTGGGCCCTCTCTGTCGACCAGGTT 61	AUTHORS	none.	TITLE	complement(join(3851..4162,4210..4629,4674..4877,
gene	1	AUTHORS	none.	TITLE	4921..5024,5071..5452))
mRNA	Db 255 CAGGACCTCCAGGACGACGACGACGACGACGACGACGACGACGACGACGACGAC 314	AUTHORS	none.	TITLE	/gene="F57B1.2"
CDS	1	AUTHORS	none.	TITLE	complement(join(3851..4162,4210..4629,4674..4877,
gene	Qy 62 TACCACATATACAGGAGAAATAGTGAA 90	AUTHORS	none.	TITLE	4921..5024,5071..5452))
mRNA	1	AUTHORS	none.	TITLE	/gene="F57B1.2"
CDS	Db 315 TTCCAGGACCTCCAGGACGACGACGACGACGACGACGACGACGACGACGACGAC 343	AUTHORS	none.	TITLE	/note="predicted using Genefinder
gene	1	AUTHORS	none.	TITLE	CDNA EST yk669g9.3 comes from this gene
mRNA	Qy 62 TACCACATATACAGGAGAAATAGTGAA 90	AUTHORS	none.	TITLE	CDNA EST yk669g9.5 comes from this gene"
CDS	1	AUTHORS	none.	TITLE	/codon_start=1
gene	Db 315 TTCCAGGACCTCCAGGACGACGACGACGACGACGACGACGACGACGACGACGAC 343	AUTHORS	none.	TITLE	/protein_id="CAB01511.1"
gene	1	AUTHORS	none.	TITLE	/db_xref="GI:387753"
mRNA	Qy 62 TACCACATATACAGGAGAAATAGTGAA 90	AUTHORS	none.	TITLE	/db_xref="SPTREMBL:Q20924"
CDS	1	AUTHORS	none.	TITLE	/translation="MALRHTLISPOFSNRHSPVTRSVTSRTGVHOPLDSTPTVTRDSO
gene	Db 315 TTCCAGGACCTCCAGGACGACGACGACGACGACGACGACGACGACGACGACGAC 343	AUTHORS	none.	TITLE	PCTITGRIQPHESADSEIDLNSSKFIYKEHFSYKEITSMKKEMWYDMLRYIRMVR
gene	1	AUTHORS	none.	TITLE	RRFVPTWAQFKRTLMAVFLFAMLYKYARDCLFDGTHHNSGSAKADKANASEKQKFH
mRNA	Qy 62 TACCACATATACAGGAGAAATAGTGAA 90	AUTHORS	none.	TITLE	QTISNLRAEFAHDKQLDFKTDHLEKLEHLSKGSKEAISAEELKQIKLWQAEISD
CDS	1	AUTHORS	none.	TITLE	ALQOMKEIDDAKSTKIITHPEKAPETAPTASLPSSQLQPMHTRALLGVNVAIS
gene	Db 255 CAGGACCTCCAGGACGACGACGACGACGACGACGACGACGACGACGACGACGAC 314	AUTHORS	none.	TITLE	LIGASIDHSCSSRPVSADGFFYDFMSYFGTFOGAYALLDDSCCTCKWQPLVANCEYDKRAT
gene	1	AUTHORS	none.	TITLE	LTVKLARFVLPKVSVOHVRSGLVFNHAPKLYDVACTDSCCTCKWQPLVANCEYDKRAT
mRNA	Qy 62 TACCACATATACAGGAGAAATAGTGAA 90	AUTHORS	none.	TITLE	DGSDDEQFCFSVPTIQNHSPINHVQFRFRHNGMDMPKTCAYLIRVGEVDPDKETO
CDS	1	AUTHORS	none.	TITLE	PMTDNGTESKLESAIVNSVSETA"
gene	Db 315 TTCCAGGACCTCCAGGACGACGACGACGACGACGACGACGACGACGACGACGAC 343	AUTHORS	none.	TITLE	complement(join(5920..6069,6126..6209,6280..6408,

PUBMED
REMARK
REFERENCE
AUTHORS
TITLE
JOURNAL

9851916
The C.elegans Sequencing Consortium.
2 (bases 1 to 269619)
Sulston,J.E.
Direct Submission
Submitted (27-OCT-1999) Nematode Sequencing Project, Sanger
Institute, Hinxton, Cambridge CB10 1SA, England and Department of
Genetics, Washington University, St. Louis, MO 63110, USA. E-mail:
jes@sanger.ac.uk or rj@nematode.wustl.edu
On May 14, 2001 this sequence version replaced gi:5730145.
Coding sequences below are predicted from computer analysis, using
predictions from Genefinder (P. Green, U. Washington), and other
available information.

COMMENT

Current sequence finishing criteria for the C. elegans genome
sequencing consortium are that all bases are either sequenced
unambiguously on both strands, or on a single strand with both a
dye primer and dye terminator reaction, from distinct subclones.
Exceptions are indicated by an explicit note.
For a graphical representation of this sequence and its analysis
see:- [http://wormbase.sanger.ac.uk/perl/ace/elegans/seq/sequence?](http://wormbase.sanger.ac.uk/perl/ace/elegans/seq/sequence?name=Y51H4A)
[name=Y51H4A](http://wormbase.sanger.ac.uk/perl/ace/elegans/seq/sequence?name=Y51H4A)

IMPORTANT: This sequence is NOT necessarily the entire insert of
the specified clone. It may be shorter because we only sequence
overlapping sections once, or longer because we arrange for a small
overlap between neighbouring submissions.
This sequence is the entire insert of clone Y51H4A. The true left
end of clone Y43D4A is at 193825 in this sequence. The start of
this sequence (1..115) overlaps with the end of sequence AL713992.
The end of this sequence (269514..269619) overlaps with the start
of sequence AL132846.

FEATURES
source

Location/Qualifiers
1..269619
/organism="Caenorhabditis elegans"
/db_xref="taxon:6239"
/chromosome="IV"
/clone="Y51H4A"
complement(join(2293..2346,2393..2610,2668..2758,
3398..3694,5250..5357))
/gene="Y51H4A.1"
complement(join(2293..2346,2393..2610,2668..2758,
3398..3694,5250..5357))
/gene="Y51H4A.1"
/note="predicted using Genefinder"
/codon_start=1
/protein_id="CAB61134.1"
/db_xref="GI:6434492"
/db_xref="SPTREMBL:O9NAE5"

gene

CDS

/translation="MLDVLLVLTILLFFYFGKLIQKIVENSIEKINSNGIS
LAENKNLEKSSNDPENAEETEMVIELHTEAPVENYLEEEIQIKVPEVEDTEKT
PAGSPKSTSSCISEILSEQSIGTCAKLYCTNRPITYGFRHTAEPVCCVLARFQFO
NENCFKSTISPRSPFLDFATSSYSAPLSNFTRETPNIVPRMSEFEFLHKCIEK
RKSDIRDLRIOSAKRRRRFSEHLQQVQDHIDY"
complement(join(12157..12228,13962..14312,16459..16593,
17111..17131))
/gene="Y51H4A.3"
complement(join(12157..12228,13962..14312,16459..16593,
17111..17131))
/gene="Y51H4A.3"
/note="contains similarity to Pfam domain: PF00071 (Ras
family), Score=317.2, E-value=6.4e-92, N=1
cDNA EST yk60d10.5 comes from this gene
cDNA EST yk103g7.5 comes from this gene
cDNA EST yk77f3.5 comes from this gene
cDNA EST yk236a7.5 comes from this gene
cDNA EST yk435f1.5 comes from this gene
cDNA EST yk531a12.5 comes from this gene
cDNA EST yk592g8.5 comes from this gene
cDNA EST yk651f5.5 comes from this gene
cDNA EST yk671f3.5 comes from this gene
cDNA EST yk565h1.5 comes from this gene"
/codon_start=1
/protein_id="CAB63379.1"
/db_xref="GI:6580310"

gene

CDS

gene
CDS

/db_xref="SWISS-PROT:Q22038"
/translation="MAAIRKKLVIVGDGAGCKTCLLIIVFSKQDPDVVYVTFVFNVA
DIEVDGKVELALWDTAGQEDYDRPLSYSDTDVILMCFSDSIDSPLENPEKWTPE
VRHCPNVPIILVGNKRDLSDPOTVRELAKMKQEPVKPEQGRAIAEQIGAFALICS
AKTKDGRVFEKATQAALQOKKKKSCMILL"
complement(join(21593..21756,22186..22384,22466..22675))
/gene="Y51H4A.2"
complement(join(21593..21756,22186..22384,22466..22675))
/gene="Y51H4A.2"
/codon_start=1
/protein_id="CAB63380.1"
/db_xref="GI:6580311"
/db_xref="SPTREMBL:O9U265"

/translation="MKSVGGSRLLLLLRRRHHISLLPLGRFPSSLFHFDPHIHY
IHGLAAPROKQKALDFKFOHSTATALCGGRVPLLYARKLSRPHMPHGAHRSS
VDNKKRRRGKAKPILTOFISKSFLLDFDKMYKNEVELSYFKSSIFCTFFPKKTIYFPR
LLGDIPAPLKLVSRAAERERAAITRIV"
complement(join(31618..31755,32295..32402,33420..33614,
33718..33834))
/gene="Y51H4A.4"
complement(join(31618..31755,32295..32402,33420..33614,
33718..33834))
/gene="Y51H4A.4"
/note="predicted using Genefinder
cDNA EST yk282h7.3 comes from this gene
cDNA EST yk340d3.3 comes from this gene
cDNA EST yk479g1.3 comes from this gene
cDNA EST yk282h7.5 comes from this gene
cDNA EST yk340d3.5 comes from this gene
cDNA EST yk479g1.5 comes from this gene
cDNA EST yk486f5.5 comes from this gene
cDNA EST yk573a7.3 comes from this gene
cDNA EST yk573a7.5 comes from this gene"
/codon_start=1
/protein_id="CAB61136.1"
/db_xref="GI:6434494"

gene

CDS

/translation="MMTDIOOLAAARMCDSDSTKAFNMVYQNPYAGAGPDKNYSAA
TALPFFFAQHEAPYANPHTDSFSSSTSTSTSSQOQSGSSKKKKPVVPENOK
DEAYLDRRRNNEAARKSRKKYDODNSVRVYLERENOCRLRVYVQOLOQNESMR
QHLLQNPGTIDTMDPIAMAMPsi"
complement(join(50748..50898,51588..51977,52741..52920,
53245..53366,53557..53611,54196..54274,54700..54854,
55009..55055,57258..57575))
/gene="Y51H4A.5"
complement(join(50748..50898,51588..51977,52741..52920,
53245..53366,53557..53611,54196..54274,54700..54854,
55009..55055,57258..57575))
/gene="Y51H4A.5"
/note="predicted using Genefinder
contains similarity to Pfam domain: PF01764 (Lipase),
Score=244.3, E-value=5.4e-70, N=1"
/codon_start=1
/protein_id="CAB61137.1"
/db_xref="GI:6434495"
/db_xref="SPTREMBL:O9NAE3"

gene

CDS

/translation="MERINCOLYLLHFLLYVCCWIVGTSALGRAEYVTCSEKHTINL
DKDATREMSDDHDKLKYRTYQWCDNTSLKYSRFOILHSNALRIEDFACNCTCFNRNRI
LFEDVKNTTKKQHNFKQKKYKTFENVFKREKTLNLELFACTNCTCFNRNRI
FNSAANLFATILVFLIFWQFFDLEASKCTDCIASGKHCWCTVNECGSPCATSIS
KSLNCPBLPKAHAYDDNFVTKIVLAASALNPENPEKCFKPRLTMTLSKKFSANCS
EVGPQSCFSTSPDTQKVLVMSFRGTDSPLOLTDLDFETGKKQFTPDAGNIFTY
FYDAFFLWAGLOODJROLKYKYDYELWYTHGSLGALASVAASYVHHGIFTSDK
VKLVLTQPTGTDYATWHDKNFPYSFRIVHRDRDLVAHIPPOGADKLFHRSEVWY
NNNSMTPEPHYCAEADGFCSSRELDTNADHLTYFGVIFPGWQTKC"
join(61826..61986,62058..62127)
/gene="Y51H4A.6"
join(61826..61986,62058..62127)
/gene="Y51H4A.6"
/codon_start=1
/protein_id="CAB63381.1"
/db_xref="GI:6580312"
/db_xref="SPTREMBL:O9U264"

gene

CDS

PUBMED		9032360	Location/Qualifiers	
FEATURES			l..3734	
source			/organism="Saaimirine herpesvirus 2"	
			/villon	
			/strain="C484"	
			/specific_host="saamiri sclureus"	
			/sub_species=" subgroup C"	
			/db_xref="taxon:10381"	
			/lab_host="Owl monkey kidney cells"	
			/notes="this strain is different from strain C484 in database entry M31964; transformation-associated region"	
gene			complement(473..1270)	
CDS			/gene="Tipc484"	
			complement(473..1270)	
			/gene="Tipc484"	
			/codon_start=1	
			/protein_id="CAA67872.1"	
			/db_xref="GI:1743283"	
			/db_xref="SPTREMBL:P88825"	
			/translation="MANEGEEIETLEFFPETEKERKEDEKLSSCSEETTDTSSSSSDSHVPATPVNVIIIONSTRTELONTKFVANECKTELGTGFOGKLSCSPETTTATSSYKSQASVCIEENGNDNETSYRPNVLTNLSLYTTFEDARAQKGMYRKVEDLSGFLEKYDPDYRRKRDLSENATWDPMGPTPALPPRPANLGERQASTVRLHVKESNCKOPKPRKANERNIVDKLRLENKVNAICTLVVVILAVLLLTVALSLIHGMKS"	
gene			complement(1588..1896)	
CDS			/gene="Stpc484"	
			complement(1588..1896)	
			/gene="Stpc484"	
			/codon_start=1	
			/protein_id="CAA67871.1"	
			/db_xref="GI:1743282"	
			/db_xref="SPTREMBL:P88824"	
			/translation="MASEPNIARYPIEETGDGRPGPPGPPOGGPCOGPPCPAGPAGP GPPGPPGPPGPCPPGPPGLGLEVTNLLGIIVLLLLLIIVALLLVSKLVVN "	
BASE COUNT	979 a	769 c	676 g	1310 t
ORIGIN				
	Query Match	33.5%;	Score 34.8;	DB 14; Length 3734;
	Best Local Similarity	62.8%;	Pred. No. 2.6;	
	Matches	54;	Conservative	0; Mismatches 32; Indels 0; Gaps 0;
QY	1	GCTGGACCTCCAGCACACCACCGCCCCCAAGAAGTGCGGCCTCTGTGTGCCACCATG 60		
Dd	1776	GCAGGACCTCCAGGACCACACGACCTCCAGGACCACCGACCTCCAGGACCACCATG 1717		
QY	61	TTACCACAATATACAGAGAAATAAG 86		
Dd	1716	CCTCCAGGACTCCAGGACCACCATG 1691		
RESULT 13				
U58736/c			5522 bp	DNA linear INV 22-MAY-2002
LOCUS				
DEFINITION			Caenorhabditis elegans cosmid EGAP7, complete sequence.	
ACCESSION			U58736	
VERSION			U58736.1	GI:1326297
KEYWORDS			HTG.	
SOURCE			Caenorhabditis elegans.	
ORGANISM			Caenorhabditis elegans	
REFERENCE			Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;	
AUTHORS			Rhabditidae; Rhabditidae; Peloderinae; Caenorhabditis.	
TITLE			1 (bases 1 to 5522)	
JOURNAL			Waterston,R.	
MEDLINE			Genome sequence of the nematode C. elegans: a platform for investigating biology. The C. elegans Sequencing Consortium	
PUBMED			Science 282 (5396), 2012-2018 (1998)	
AUTHORS			99059613	
TITLE			9851916	
			2 (bases 1 to 5522)	
			Miller,N.	
			The sequence of C. elegans cosmid EGAP7	

Db 17444 GCTGGATCCCGACGACCGCGGACATCCAGGAACGCCGGAGCTCCAGGAACCCCCAGGA 17385

Qy 61 TTACCACAATATACAGGAG 79
 |||| | || |||||

Db 17384 GCACCGAATCCAGGAG 17366

RESULT 15
 HSC3A1R 3234 bp mRNA linear PRI 05-AUG-1995
 LOCUS Human COL3A1 mRNA for pro alpha-1 (III) collagen.
 DEFINITION X15332
 ACCESSION X15332
 VERSION X15332.1 GI:29545
 KEYWORDS COL3A1 gene; collagen.
 SOURCE Homo sapiens.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 3234)
 AUTHORS Janeczko,R. and Ramirez,F.
 TITLE Direct Submission
 JOURNAL Submitted (19-MAY-1989) Janeczko R., Ramirez F., Suny Health
 Science Centre, 450 Clarkson Avenue - Box 44, Brooklyn NY 11203, U
 S A
 2 (bases 1 to 3234)
 AUTHORS Janeczko,R.A. and Ramirez,F.
 TITLE Nucleotide and amino acid sequences of the entire human alpha 1
 (III) collagen
 JOURNAL Nucleic Acids Res. 17 (16), 6742 (1989)
 MEDLINE 89386015
 PUBMED 2780304
 COMMENT The sequence overlaps with that reported by Chu et. al. in J. Biol.
 Chem. 260:4357-4363(1985), by Toman et. al. in Nucl. Acids Res.
 16:7201-7201(1988) and by Mankoo et. al. in Nucl. Acids Res.
 16:2337-2337(1988).

FEATURES
 source Location/Qualifiers
 1..3234
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /map="2q31"
 <1..>3234
 /codon_start=1
 /product="alpha-1 (III) collagen"
 /protein_id="CAA33387.1"
 /db_xref="GI:930045"
 /db_xref="SWISS-PROT:P02461"
 /translation="QNYSPQYSDYDKSGGVAVGLAGYPCGAPGCPGPGPGTSGHP
 GSPGPGCPGPGCGAGGPPGPPGPAIGSPGAKGDSGRGPRGPRGDPGLPGPPG
 IKGPAIGYCPGPMKHGRGDFRNGKEGTAGPLKGELNGPAGGPPGMPGRCGPE
 RGRPLPAGAGAGNDGARGSDQPPGPGTAGFPSPGAKGVGPPGAGSPGSGNAG
 GORGEPPGKHAGACQPPGPGINGSPGKGEMGPAGIPGAPGLMGAGPPGPPGAG
 APGLRGAGEPKNGAKPPGPRGERGEAGIPGVAKGEDGKSGSPDGPAGNLPGA
 AGERGALSRGPAGNGIIGEGKGPAGERGAPGAPRGAAAGEPRGDPGPGMRGMP
 GSPGPGSDCKPPGPGQSGESGRPPGSPGPGQPGVPGPGRKNDGAPGKNGRGG
 GPPGPGPPGPKNGISYGGQPPGTPGPGDKDGTDPGPGQLGTPGTGGTPEENG
 PGPSPGKGAGAPGAPGKGADGACGERGPPGLAGAPGLRGAGPPGEGKGAAGP
 PGPAAGTGLQCMGREGGLGSPGPKDGPPGAGIPGVAKGEDGKSGSPDGPAGNLPGA
 PAGACDKGEGAGPLIPLAGPSPGERGENTGPPGAPGAPGQNGEPGCGKGERGA
 PGEKGGPPGVAGPPGSGSPAGPPGQGVKGERGSPGPGAAAGFPGARGLPPGGSN
 GNPDPGSPGPKDPPGPPAGNTGAPSPGVSGPKDAGOPGKSGSPGAGQGPAGP
 PLGIATIGTGRLAGPPGPPRSPGQVKGESKPGANGLSGRGPPGPGPLGCL
 ACTAGSPGDRGLDPLGRDGSPPGKDRGNGSPGAPGAPGHPGPPGPPGPAKGS
 GDRGESGPAGPACGAPGAPGQGRGKETGERGAAGIKGHRGPPGPPGAPG
 SPGPAQCAIGSPGAPGPPGVPSPGPKDGTSGHPGIPGPPGPRNGRGERGSEGS
 PPHGPGSPGPPGPAAGCGGVAIAIGAEKAGGAFYIYGDPM"

BASE COUNT 664 a 861 c 1106 g 603 t

ORIGIN

Query Match 32.7%; Score 34; DB 9; Length 3234;
 Best Local Similarity 61.1%; Pred. No. 4.5;
 Matches 55; Conservative 0; Mismatches 35; Indels 0; Gaps 0;

Qy 1 GCTGGACCTCCAGCACACCCAGGCCGCCCAAGAAAGTGGGGCCTCTCTGTGTGACACCAAGT 60

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 19, 2003, 23:48:42 ; Search time 1130.07 Seconds
(without alignments)
6406.126 Million cell updates/sec

Title: us-09-997-610-1_copy_2_448

Perfect score: 447

Sequence: 1 atagtggtcatacctgtctt.....cttggtgtctgttaaggca 447

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*

1: em_estba:*

2: em_esthum:*

3: em_estlin:*

4: em_estm:*

5: em_estov:*

6: em_estpl:*

7: em_estro:*

8: em_htc:*

9: gb_est1:*

10: gb_est2:*

11: gb_htc:*

12: gb_est3:*

13: gb_est4:*

14: gb_est5:*

15: em_estfun:*

16: em_estom:*

17: gb_gss:*

18: em_gss_hum:*

19: em_gss_inv:*

20: em_gss_pln:*

21: em_gss_vrt:*

22: em_gss_fun:*

23: em_gss_mam:*

24: em_gss_mus:*

25: em_gss_other:*

26: em_gss_pro:*

27: em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	113.6	25.4	499	14	BM967732
2	76.4	17.1	504	12	BE757275
3	76.4	17.1	526	12	BE757276
4	61.8	13.8	612	17	AG088117
5	59.6	13.3	538	9	AA777621
6	58.6	13.1	440	10	BE062167

c	7	58.6	13.1	475	17	B36584	B36584 HS-1041-A1-
c	8	58.6	13.1	537	10	BE079777	BE079777 RC6-BT062
c	9	58.6	13.1	563	10	BE079876	BE079876 RC6-BT062
c	10	58.6	13.1	678	10	AW813783	AW813783 RC3-ST019
c	11	58.6	13.1	727	10	AV731140	AV731140 AV731140
c	12	58.6	13.1	2615	10	BE420422	BE420422 32-393 hu
c	13	58	13.0	342	17	AO012504	AO012504 CIT-HSP-2
c	14	58	13.0	345	12	BF881529	BF881529 QW1-ET018
c	15	58	13.0	670	17	AG039990	AG039990 Pan trogl
c	16	57.8	12.9	397	13	BI180953	BI180953 T4F01 he
c	17	57	12.8	425	10	AW517269	AW517269 xq08049.x
c	18	57	12.8	426	9	AA372508	AA372508 EST84499
c	19	57	12.8	660	17	AG019007	AG019007 Homo sapi
c	20	57	12.8	666	17	AG141320	AG141320 Pan trogl
c	21	57	12.8	683	17	AG019080	AG019080 Homo sapi
c	22	57	12.8	1036	13	BM462674	BM462674 AGENCOURT
c	23	56.8	12.7	577	10	AV716883	AV716883 AV716883
c	24	56.6	12.7	662	10	BE390081	BE390081 601285704
c	25	56.4	12.6	284	9	AA360873	AA360873 EST70071
c	26	56.4	12.6	461	12	BF828364	BF828364 MRL-HN006
c	27	56.4	12.6	461	12	BF829000	BF829000 MRL-HN006
c	28	56.4	12.6	545	10	BE279213	BE279213 601156706
c	29	56.4	12.6	609	10	BE389158	BE389158 601285954
c	30	56.4	12.6	667	17	AG080755	AG080755 Pan trogl
c	31	56.4	12.6	668	17	AG080765	AG080765 Pan trogl
c	32	56.4	12.6	698	12	BE729966	BE729966 601562450
c	33	56.4	12.6	705	17	AG052122	AG052122 Pan trogl
c	34	56.4	12.6	721	12	BE728927	BE728927 601562364
c	35	56.4	12.6	761	12	BE728912	BE728912 601562343
c	36	56.4	12.6	795	12	BF025976	BF025976 601669973
c	37	56.4	12.6	800	12	BF203746	BF203746 601868725
c	38	56.4	12.6	885	12	BF304683	BF304683 601888096
c	39	56.4	12.6	1080	14	BQ422247	BQ422247 AGENCOURT
c	40	56	12.5	315	9	AA719635	AA719635 z952e04.s
c	41	56	12.5	645	17	AG014562	AG014562 Homo sapi
c	42	55.8	12.5	360	10	AW898344	AW898344 RC3-NN007
c	43	55.4	12.4	364	17	AQ665181	AQ665181 HS_5344_B
c	44	55.4	12.4	388	17	B78843	B78843 CIT-HSP-734
c	45	55.4	12.4	453	17	AQ393565	AQ393565 CITBI-E1-

ALIGNMENTS

RESULT 1	BM967732	BM967732	499 bp	mrna	linear	EST 20-MAR-2002
LOCUS	LM24HW0134	Bos taurus	LM-24-HW	cdna	library	Bos taurus cdna clone
DEFINITION	LM-24-HW-011-34 (5'), mRNA sequence.					
ACCESSION	BM967732					
VERSION	BM967732.1	GI:19561919				
KEYWORDS	EST.					
SOURCE	COW.					
ORGANISM	Bos taurus					
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovinae; Bos.					
AUTHORS	Yoon,D.H., Jang,Y.S., Kim,T.H., Park,E.W., Lee,H.K., Chung,E.R., Sun,S.S. and Cheong,I.C.					
TITLE	Gene Expression Profiling of the Bovine skeletal muscle					
JOURNAL	Unpublished (2002)					
COMMENT	Contact: Dr. Du-Hak Yoon National Livestock Research Institute, RDA 564 Omoekchun-dong, Suwon, 441-350, Korea Tel: 82 31 290 1593 Fax: 82 31 290 1792 Email: dhyoon@rda.go.kr Insert Length: 499 Std Error: 0.00 Seq primer: CAGGAACAGCTATGAC POLYA-No.					
FEATURES	Location/Qualifiers					
source	1..499					

/organism="Bos taurus"
/db_xref="taxon:9913"
/clone="LM-24-HW-011-34 (5')"
/clone.lib="Bos taurus LM-24-HW cDNA library"
/sex="six males mixed"
/tissue_type="longissimus dorsi"
/cell_type="myocyte"
/dev_stage="24 months old"
/lab_host="XLI-BlueMRF'strain"
/note="Organ: skeletal muscle; Vector: Uni-ZAPXR; Site_1:
EcoRI; Site_2: Xho I"
BASE COUNT 129 a 139 c 131 g 100 t
ORIGIN

Query Match 25.4%; Score 113.6; DB 14; Length 499;
Best Local Similarity 64.2%; Pred. No. 4.4e-21;
Matches 203; Conservative 0; Mismatches 109; Indels 4; Gaps 2;

QY 46 GAAGTTGCTGGACCTCCAGCACACCCAGGCCGCCAGAGAGTGGGGCTCTCTGTGCA 105
Db 174 GGACCTCCGGGCCACAGGACACCCAGGCGCTCCAGGTATAAGAGGGCCTCCAGGTATA 233
QY 106 CCAGGTTTACCATATATACAGGAG---AAATAAGTGAATGACAAAATGCCCTGTCT 162
Db 234 AGAGGAATACCAAGTTTGCCAGGTCCGGGGAACCTCCAGGACCAAGTGTAAATGCCCA 293
QY 163 GATATAGAAAGTGCAGCCTTTACTGTGAAGCTCAGTGGAAACTTCCCTCTCTTCAG 222
Db 294 TCCACAGACAGTCGCGCTTCTACTGTGAAGCTCAGTGGCCAGTTGCCCTTCCCTTCAAAG 353
QY 223 CCCAT-CATCTTCACAGGGGTCTGTACAATGCCAGAGGGATTTAAAGGAGGCCATGG 281
Db 354 CCGTGGCCCTTTCACAGAGGTCTGTACAATGCCAGAGAGACTTACAGGAGACACTGG 413
QY 282 AGCTTTGCTTCAGGGTGCCTGGGAATTAATCTACCTCCAGCTTTGATGTTAGCTGCATCA 341
Db 414 GGTCTTCACATCAGGGTGCCAGGAATTACCATTTCCTTCTATGTGGATCTCCATCA 473

QY 342 TTGCAAGGTGAATATT 357
Db 474 CTGCAAGGTGACTGTT 489

RESULT 2
BE757275
LOCUS BE757275 504 bp mRNA linear EST 25-APR-2001
DEFINITION 211668 MARC 2BOV Bos taurus cDNA 5', mRNA sequence.
ACCESSION BE757275
VERSION BE757275.1 GI:10171267
KEYWORDS EST.
SOURCE COW.
ORGANISM Bos taurus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovidae; Bovinae; Bos.
1 (bases 1 to 504)
Smith,T.P.L., Grosse,W.M., Freking,B.A., Roberts,A.J., Stone,R.T.,
Casas,E., Wray,J.E., White,J., Cho,J., Fahrnkruug,S.C., Bennett
,G.L., Heaton,M.P., Laegreid,W.W., Rohrer,G.A., Chitko-Mckown,C.G.,
Perteau,G., Holt,I., Karamycheva,S., Liang,F., Quackenbush,J. and
Keele,J.W.

TITLE
Sequence evaluation of four pooled-tissue normalized bovine cDNA
libraries and construction of a gene index for cattle
JOURNAL
MEDLINE
COMMENT
Genome Res. 11 (4), 626-630 (2001)
Contact: Smith TPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390
Email: smithemail.marc.usda.gov
v0.980904.e. Vector identified by cross_match with the -minscore 18

and -minmatch 12 options.
PCR Primers
FORWARD: AGGAAACAGCTATGACCAT
BACKWARD: GTTTCCAGTCAGCAGC
Plate: 63 row: G column: 13
Seq primer: ATTTAGGTGACACTATAG.
FEATURES
Location/Qualifiers
1..504
/organism="Bos taurus"
/db_xref="taxon:9913"
/clone.lib="MARC 2BOV"
/tissue_type="pooled"
/lab_host="DHI10B"
/note="Vector: pCMV SPORT6; Site_1: NotI; Site_2: SalI;
Library made from pooled tissue from testis, thymus,
semiteendonosus muscle, longissimus muscle, pancreas,
adrenal, and endometrium."
BASE COUNT 99 a 131 c 146 g 128 t
ORIGIN

Query Match 17.1%; Score 76.4; DB 12; Length 504;
Best Local Similarity 61.6%; Pred. No. 1.2e-10;
Matches 122; Conservative 0; Mismatches 76; Indels 0; Gaps 0;

QY 174 GTCAGCCTTTACTGTGAAGCTCAGTGGAAACTTCTCTCTTCAAGCCCATCATCTT 233
Db 303 GTCTGCCCTTTCCGTCAAGCTCAGTGGGCTTTCCAGGACCTCCAGCCCATTTGCTT 362
QY 234 CACAGGGTCTCTGTACAATGCCAGAGGGATTTAAAGGAGGCCATGGGAGTCTTTGCTT 293
Db 363 CCAGGAAGTTCTGTACAACCATCAGGGCCACTTCGACCCGCCACTGGTGTCTTCAAGTG 422
QY 294 CAGGAGTCTGGGAATTAATCTACTCTCAGCTTTGATGTTAGCTGCATCATTCGAAGTGAA 353
Db 423 CAGCCTCCCTGGTGTGACCACTTTGGCTTTGACATTTGAGTTGTTTCAGAGTGTCTCAA 482

QY 354 TATTTGGCTAATGAGGAA 371
Db 483 GGTGGTCTAATGCGAA 500

RESULT 3
BE757276
LOCUS BE757276 526 bp mRNA linear EST 25-APR-2001
DEFINITION 211669 MARC 2BOV Bos taurus cDNA 5', mRNA sequence.
ACCESSION BE757276
VERSION BE757276.1 GI:10171268
KEYWORDS EST.
SOURCE COW.
ORGANISM Bos taurus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovidae; Bovinae; Bos.
1 (bases 1 to 526)
Smith,T.P.L., Grosse,W.M., Freking,B.A., Roberts,A.J., Stone,R.T.,
Casas,E., Wray,J.E., White,J., Cho,J., Fahrnkruug,S.C., Bennett
,G.L., Heaton,M.P., Laegreid,W.W., Rohrer,G.A., Chitko-Mckown,C.G.,
Perteau,G., Holt,I., Karamycheva,S., Liang,F., Quackenbush,J. and
Keele,J.W.

TITLE
Sequence evaluation of four pooled-tissue normalized bovine cDNA
libraries and construction of a gene index for cattle
JOURNAL
MEDLINE
COMMENT
Genome Res. 11 (4), 626-630 (2001)
Contact: Smith TPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390
Email: smithemail.marc.usda.gov
v0.980904.e. Vector identified by cross_match with the -minscore 18

PCR Primers
FORWARD: AGGAAACAGCTATGACCAT
BACKWARD: GTTTCCAGTCAGCAGC
Plate: 63 row: G column: 13
Seq primer: ATTTAGGTGACACTATAG.
FEATURES
Location/Qualifiers
1..504
/organism="Bos taurus"
/db_xref="taxon:9913"
/clone.lib="MARC 2BOV"
/tissue_type="pooled"
/lab_host="DHI10B"
/note="Vector: pCMV SPORT6; Site_1: NotI; Site_2: SalI;
Library made from pooled tissue from testis, thymus,
semiteendonosus muscle, longissimus muscle, pancreas,
adrenal, and endometrium."
BASE COUNT 99 a 131 c 146 g 128 t
ORIGIN

FORWARD: AGGAACAGCTATGACCAT
 BACKWARD: GTTTCCTCCAGTCACGACG
 Plate: 63 row: G column: 14
 Seq primer: ATTAGTGACACTATAG.

FEATURES

source
 Location/Qualifiers
 1..526
 /organism="Bos taurus"
 /db_xref="taxon:9913"
 /clone_lib="MARC 2BOV"
 /tissue_type="pooled"
 /lab_host="DH10B"

/note="Vector: pCMV SPORT6; Site_1: NotI; Site_2: SalI;
 Library made from pooled tissue from testis, thymus,
 semitendinosus muscle, longissimus muscle, pancreas,
 adrenal, and endometrium."

BASE COUNT 106 a 137 c 153 g 130 t

ORIGIN

Query Match 17.1%; Score 76.4; DB 12; Length 526;
 Best Local Similarity 61.6%; Pred. No. 1.2e-10;
 Matches 122; Conservative 0; Mismatches 76; Indels 0; Gaps 0;

QY 174 GTCAGCCTTACTCTGAGCTCACTGAAACCTTCCTCTTCCCTTCAAGCCCATCATCTT 233

Db 303 GTCTGCCCTTTCCGTCAGAGCTGAGTGGGCTTTCCAGGACCCCTCCAGCCCATTTGCTT 362

QY 234 CACAGGGGCTCTTACAAATCCAGAGGGATTTAAAGGAGGCCATGGGAGCTTTTGCTTG 293

Db 363 CCAGGAAGTCTGTACAAACCATCAGGCCACTTCGACCCCGCCACTGGTGTTCAGCTG 422

QY 294 CAGGGTCCCTGGGAATTTACTACCTCCAGCTTTGATTTGAGCTGCATCATTTGCAAGGTGAA 353

Db 423 CAGCGTCCCTGGTGTGTACCACTTTGGCTTTGACATTTGAGTTGTTTCAGAGTGTCTGCAA 482

QY 354 TATTGGCTAATGAGGAA 371

Db 483 GGTGGGTCTAATCGGAA 500

RESULT 4

AG088117

LOCUS Pan troglodytes DNA, clone: PTB-086P15.F, genomic survey sequence.
 DEFINITION AG088117

ACCESSION AG088117.1 GI:16639919

KEYWORDS

SOURCE

ORGANISM Pan troglodytes male lymphoblast DNA, clone_lib:PTB Chimpanzee male
 BAC Library clone:PTB-086P15.F.

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.

1

Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T. D., Yada, T.,

Totoki, Y., Watanabe, H. and Sakaki, Y.

BAC end sequences of Library PTB

Unpublished

2 (bases 1 to 612)

Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T. D., Yada, T.,

Totoki, Y., Watanabe, H. and Sakaki, Y.

Direct Submission

Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical

and Chemical Research (RIKEN), Genomic Sciences Center (GSC);

1-7-22 Suehiro-chou, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan

(E-mail: chimbes@gsc.riken.go.jp, URL: http://hgp.gsc.riken.go.jp/,

Tel: 81-45-503-9111, Fax: 81-45-503-9170)

Clones are derived from the chimpanzee BAC library PTB This BAC end

was generated during the R&D process and may have higher chance of

clone tracking errors.

PRIMERS

Sequencing: -21M13

LIBRARY

Vector : pKS145

R.Site 1 : SacI

R.Site 2 : SacI.

Location/Qualifiers

1..612

/organism="Pan troglodytes"

/db_xref="taxon:9598"

/clone="PTB-086P15.F"

/sex="male"

/cell_type="lymphoblast"

/clone_lib="PTB Chimpanzee Male BAC Library"

BASE COUNT 189 a 111 c 163 g 149 t

ORIGIN

Query Match 13.8%; Score 61.8; DB 17; Length 612;

Best Local Similarity 77.3%; Pred. No. 1.7e-06;

Matches 75; Conservative 0; Mismatches 22; Indels 0; Gaps 0;

QY 351 GAATATTTGGCTAATGAGGAGCAAAATTTGGCTTAATAAGCAAAATTTCTTAAGCAGCA 410

Db 269 GAACCTTTGAACCTTGAGAGAGATCATTTAGGCTATCAGGAGGAGAAATTTCTTAAGCAGCA 328

QY 411 AAGCATTTCAAGAGTGTGACTTGGGTGCTGTTAAAGGCA 447

Db 329 AAGCATTTCAAGAGTGTGACTTGGGTGCTGTTAAAGGCA 365

RESULT 5

AA777621/c

LOCUS

DEFINITION

2195407.sl Soares_fetal_liver_spleen_INFLS_S1 Homo sapiens cDNA

clone IMAGE:448500 3' similar to contains THR.t3 THR repetitive

element ;, mRNA sequence.

ACCESSION AA777621

VERSION AA777621

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 538)

AUTHORS Hillier, L., Allen, M., Bowles, L., Dubuque, T., Geisel, G., Jost, S.,

Krizman, D., Kucaba, T., Lacy, M., Le, N., Lannon, G., Marra, M., Martin

, J., Moore, B., Scheilenberg, K., Steptoe, M., Tan, F., Theising, B.,

White, Y., Wylie, T., Waterston, R. and Wilson, R.

WashU-NCI human EST Project

Unpublished (1997)

CONTACT: Wilson RK

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: est@watson.wustl.edu

This clone is available royalty-free through LLNL ; contact the

IMAGE Consortium (info@image.llnl.gov) for further information.

Seq primer: -40m13 fwd. Et from Amersham

High quality sequence stop: 515.

Location/Qualifiers

1..538

/organism="Homo sapiens"

/db_xref="GDB:1352757"

/db_xref="taxon:9606"

/clone="IMAGE:448500"

/clone_lib="Soares_fetal_liver_spleen_INFLS_S1"

/sex="male"

/dev_stage="20 week-post conception fetus"

/lab_host="DH10B (ampicillin resistant)"

/note="Organ: Liver and Spleen; Vector: p77T3D (Pharmacia)

with a modified polylinker; Site_1: Pac I; Site_2: Eco RI;

This is a subtracted version of the original Soares fetal

liver spleen INFLS library. 1st strand cDNA was primed

with a Pac I - oligo(dT) primer [5',

AACTGGAGAAATTAATGAGATCTTTTTTTTTTTTTTTTTTTT 3'],

double-stranded cDNA was ligated to Eco RI adaptors

(Pharmacia), digested with Pac I and cloned into the Pac I


```
Db 553 AAGCATTCAAGAGGTAACCTTGGGTGCTGTTAAAGGCA 589

RESULT 13
AQ012504/c
LOCUS
DEFINITION CIT-HSP-2298M4.TF CIT-HSP Homo sapiens genomic clone 2298M4, DNA
sequence.
ACCESSION AQ012504
VERSION AQ012504.1 GI:3185069
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 342)
AUTHORS Adams,M.D., Rounsley,S.D., Zhao,S., Field,C.E., Bass,S., Linher,K.,
Golden,K., Berry,K., Granger,D., Suh,E., Wible,C., Shizuya,H.,
Simon,M. and Venter,J.C.
TITLE Use of a random BAC End Sequence Database for Sequence-Ready Map
Building (1998)
JOURNAL Unpublished (1998)
COMMENT Contact: Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: mdams@tigr.org
Clones are available from Research Genetics (info@resgen.com). BAC
end search page:
http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.
Seq primer: M13-21
Class: BAC ends.

FEATURES
source
1..342
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="ET0181"
/dev_stage="Adult"
/notes="Organ: lung_tumor; Vector: puc18; Site_1: SmaI;
Site_2: SmaI; A mini-library was made by cloning products
derived from ORESTES PCR (U.S. Letters Patent application
No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the pUC 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."
BASE COUNT 86 a 96 c 57 g 106 t
ORIGIN

Query Match 13.0%; Score 58; DB 12; Length 345;
Best Local Similarity 100.0%; Pred. No. 1.5e-05;
Matches 58; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 390 GGAAGAAATTTCTAAGCAGCAAGCATTCAAGAGTGACCTGGGTGCTGTTAAAGGCA 447
|||||
Db 194 GGAAGAAATTTCTAAGCAGCAAGCATTCAAGAGTGACCTGGGTGCTGTTAAAGGCA 137
|||||

.. RESULT 15
AG093990/c
LOCUS
DEFINITION Pan troglodytes DNA, clone: PTB-094J23.R, genomic survey sequence.
ACCESSION AG093990
VERSION AG093990.1 GI:16645792
KEYWORDS GSS.
SOURCE Pan troglodytes male lymphoblast DNA, clone_lib:PTB Chimpanzee Male
BAC Library clone:PTB-094J23.R.
ORGANISM Pan troglodytes
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Pan.
REFERENCE 1
AUTHORS Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE BAC end sequences of Library PTB
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 670)
AUTHORS Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE Direct Submission
JOURNAL Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
1-7-22 Suehiro-cho,Tsukumi-ku, Yokohama, Kanagawa 230-0045, Japan
(E-mail:chmhpbes@sc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
Tel:81-45-503-9111, Fax:81-45-503-9170)
Clones are derived from the chimpanzee BAC library PTB This BAC end
was generated during the R&D process and may have higher chance of

Db 553 AAGCATTCAAGAGGTAACCTTGGGTGCTGTTAAAGGCA 589

RESULT 14
BF881529/c
LOCUS
DEFINITION QV1-ET0181-031200-546-f03 ET0181 Homo sapiens cDNA, mRNA sequence.
ACCESSION BF881529
VERSION BF881529.1 GI:12271655
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 345)
AUTHORS Dias Neto,E., Garcia Correa,R., Verjovsky-Almeida,S., Briones,M.R.,
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
```

clone tracking errors.

PRIMERS

Sequencing: M13Rev

LIBRARY

Vector : pKSL45

R.Site 1 : SacI

R.Site 2 : SacI

Location/Qualifiers

1. .670

/organism="Pan troglodytes"

/db_xref="taxon:9598"

/clone="PTB-094J23.R"

/sex="male"

/cell_type="lymphoblast"

/clone_lib="PTB Chimpanzee Male BAC Library"

155 a 186 c 124 g 205 t

ORIGIN

BASE COUNT

Query Match 13.0%; Score 58; DB 17; Length 670;

Best Local Similarity 100.0%; Pred. No. 2e-05;

Matches 58; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 390 GGAAGAAATTTCTAAGCAGCAAGCATTCAAGAGGTGACTTGGGTGCTGTTAAAGGCA 447

Db 404 GGAGAAATTTCTAAGCAGCAAGCATTCAAGAGGTGACTTGGGTGCTGTTAAAGGCA 347

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

Search completed: February 20, 2003, 06:14:39

Job time : 1136.07 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 19, 2003, 18:01:21 ; Search time 35.9695 seconds
(without alignments)
6329.459 Million cell updates/sec

Title: us-09-997-610-1_copy_2_448

Perfect score: 447

Sequence: 1 atagtggtcatacctgtctt.....cttgggtgctgtaaaaggca 447

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 424239 seqs, 254661826 residues

Total number of hits satisfying chosen parameters: 848478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published_Applications_NA:*

- 1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
- 2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
- 3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
- 4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
- 5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
- 6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
- 7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
- 8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
- 9: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
- 10: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
- 11: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
- 12: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
- 13: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
- 14: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	447	100.0	1381	9	US-09-997-610-1
2	447	100.0	1731	9	US-09-997-610-5
3	324.2	72.5	1377	9	US-09-997-610-3
4	324.2	72.5	1731	9	US-09-997-610-7
5	114.8	25.7	425	10	US-09-960-352-15057
6	93.6	20.9	389	10	US-09-960-352-9137
7	82.8	18.5	392	10	US-09-960-352-220
8	80.2	17.9	410	10	US-09-960-352-1786
9	76	17.0	374	10	US-09-960-352-11516
10	73.2	16.4	447	10	US-09-960-352-3684
11	60.6	13.6	467	10	US-09-960-352-4241
12	58.4	13.1	273	10	US-09-864-761-25080
13	58.4	13.1	519	10	US-09-864-761-8349
14	57	12.8	1946	10	US-09-864-761-2925
15	57	12.8	25603	9	US-09-819-607-3
c 16	57	12.5	505	10	US-09-864-761-7075
17	56	12.4	520	10	US-09-864-761-8643
18	55.4	12.4	542	10	US-09-864-761-8938
c 19	55.4	12.4	542	10	US-09-864-761-8938

c 20	55.4	12.4	170834	10	US-09-835-232-7	Sequence 7, Appli
c 21	55.4	12.4	170834	10	US-09-835-232-7	Sequence 7, Appli
c 22	54.8	12.3	525	10	US-09-864-761-8932	Sequence 8932, Ap
c 23	53.8	12.0	444	10	US-09-864-761-24780	Sequence 24780, A
c 24	53.8	12.0	484	10	US-09-864-761-8387	Sequence 83, Appl
c 25	53.8	12.0	523	10	US-09-864-761-8787	Sequence 8787, Ap
c 26	53.8	12.0	543	10	US-09-864-761-8034	Sequence 8034, Ap
c 27	53.8	12.0	544	10	US-09-864-761-9454	Sequence 9454, Ap
c 28	53.2	11.9	531	10	US-09-864-761-14775	Sequence 14775, A
c 29	53.2	11.9	576	10	US-09-864-761-9294	Sequence 9294, Ap
c 30	53.2	11.9	580	10	US-09-864-761-9802	Sequence 9802, Ap
c 31	52.6	11.8	552	10	US-09-864-761-8495	Sequence 8495, Ap
c 32	52.2	11.7	220	10	US-09-864-761-33019	Sequence 33019, A
c 33	52.2	11.7	516	10	US-09-864-761-7399	Sequence 7399, Ap
c 34	52.2	11.7	520	10	US-09-864-761-8449	Sequence 8449, Ap
c 35	52.2	11.7	571	10	US-09-864-761-16494	Sequence 16494, A
c 36	52.2	11.7	600	10	US-09-864-761-7441	Sequence 7441, Ap
c 37	52.2	11.7	202001	10	US-09-734-674-3	Sequence 3, Appli
c 38	51.6	11.5	473	10	US-09-864-761-2556	Sequence 2556, Ap
c 39	51.6	11.5	3816	10	US-09-880-107-2298	Sequence 2298, Ap
c 40	51.6	11.5	42999	10	US-09-740-029-3	Sequence 3, Appli
c 41	51.6	11.5	155074	9	US-10-026-188-6	Sequence 6, Appli
c 42	51.6	11.5	180557	12	US-10-003-806-6	Sequence 6, Appli
c 43	51.6	11.5	180557	12	US-10-003-806-9	Sequence 9, Appli
c 44	51.6	11.5	148567	9	US-10-254-869-3	Sequence 3, Appli
c 45	51.2	11.5	148567	10	US-09-801-876B-3	Sequence 3, Appli

ALIGNMENTS

RESULT 1
US-09-997-610-1
; Sequence 1, Application US/09997610
; Patent No. US20020156244A1
; GENERAL INFORMATION:
; APPLICANT: Fox, Brian
; APPLICANT: Holloway, James L.
; TITLE OF INVENTION: ADIPOCYTE COMPLIMENT RELATED PROTEIN
; TITLE OF INVENTION: ZACR13
; FILE REFERENCE: 00-96
; CURRENT APPLICATION NUMBER: US/09/997,610
; CURRENT FILING DATE: 2001-11-29
; PRIOR APPLICATION NUMBER: US 60/253,924
; PRIOR FILING DATE: 2000-11-29
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: FastSeq for Windows version 3.0
; SEQ ID NO 1
; LENGTH: 1381
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (2)...(1381)
US-09-997-610-1

Query Match	100.0%	Score 447:	DB 9:	Length 1381:
Best Local Similarity	100.0%	Pred. No.	3.1e-130:	
Matches 447:	Conservative	0:	Mismatches	0:
			Indels	0:
			Gaps	0:
QY	1	ATAGTGGTGCATACCTGTCTTAATAACGGCAGTCATTAGCATGTAGCAAGTTGCTGGACCT	60	
Db	2	ATAGTGGTGCATACCTGTCTTAATAACGGCAGTCATTAGCATGTAGCAAGTTGCTGGACCT	61	
QY	61	CCAGCACACCCCGAGGCCCGCCCAAGGAGTGGGGCTCTCTGGTGCACCGAGTTTACACAA	120	
Db	62	CCAGCACACCCCGAGGCCCGCCCAAGGAGTGGGGCTCTCTGGTGCACCGAGTTTACACAA	121	
QY	121	TATACAGGAGAATAGTGAATGACAAATGCCCTGTCTCTGATATACAAAGGTCAGCC	180	
Db	122	TATACAGGAGAATAGTGAATGACAAATGCCCTGTCTCTGATATACAAAGGTCAGCC	181	
QY	181	TTTACTGTGAAGTCTAGTGGAAACTTCTCTTCTTCAAGCCCATCATCTTACAGGG	240	

```
Db 182 TTTACTGTGAAGCTCAGTGGAACCTTCCTCTCCCTTTCAAGCCCATCATCTTCACAGGG 241
QY 241 GTCCTGTACAATGCCAGAGGGATTTAAAGAGGCCATGGAGCTCTTTGCTTGCAGGGTG 300
Db 242 GTCCTGTACAATGCCAGAGGGATTTAAAGAGGCCATGGAGCTCTTTGCTTGCAGGGTG 301
QY 301 CTTGGGAATTTACTCTCCAGCTTTTGATGTGTAGCTGCATCATTTGCAAGGTGAATATTTGG 360
Db 302 CTTGGGAATTTACTCTCCAGCTTTTGATGTGTAGCTGCATCATTTGCAAGGTGAATATTTGG 361
QY 361 CTAATGAGGAAGCAAAATTTTGGCTAATAAGGAAGAAATTTCTAAGCAGCAAAAGCATTTCAA 420
Db 362 CTAATGAGGAAGCAAAATTTTGGCTAATAAGGAAGAAATTTCTAAGCAGCAAAAGCATTTCAA 421
QY 421 GAGTGACTTGGTGCTGTTAAAGGCA 447
Db 422 GAGTGACTTGGTGCTGTTAAAGGCA 448

RESULT 2
US-09-997-610-5
; Sequence 5, Application US/09997610
; Patent No. US20020156244A1
; GENERAL INFORMATION:
; APPLICANT: Fox, Brian
; APPLICANT: Holloway, James L.
; TITLE OF INVENTION: ADIPOCYTE COMPLIMENT RELATED PROTEIN
; FILE REFERENCE: 00-96
; CURRENT APPLICATION NUMBER: US/09/997,610
; PRIOR FILING DATE: 2001-11-29
; PRIOR APPLICATION NUMBER: US 60/253,924
; PRIOR FILING DATE: 2000-11-29
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 5
; LENGTH: 1731
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1731)
US-09-997-610-5

Query Match 100.0%; Score 447; DB 9; Length 1731;
Best Local Similarity 100.0%; Pred. No. 3.5e-130; Mismatches 0; Indels 0; Gaps 0;
Matches 447; Conservative 0;

QY 1 ATAGTGGTCATACCTGCTCTTAATAACGGCAGTCATTTGAGCATGTAGAAGTTGCTGGACCT 60
Db 355 ATAGTGGTCATACCTGCTCTTAATAACGGCAGTCATTTGAGCATGTAGAAGTTGCTGGACCT 414
QY 61 CCAGCACACCCAGGCCCCAGAGAAGTGGGGCCCTCCTGGTGCCACAGGTTTACCACAA 120
Db 415 CCAGCACACCCAGGCCCCAGAGAAGTGGGGCCCTCCTGGTGCCACAGGTTTACCACAA 474
QY 121 TATACAGGAGAAATAAGTGAATGACAAATGCCCTGCTCCATATAGAAAGGTCAGCC 180
Db 475 TATACAGGAGAAATAAGTGAATGACAAATGCCCTGCTCCATATAGAAAGGTCAGCC 534
QY 181 TTTACTGTGAAGCTCAGTGGAAACCTTCCTCTTCCTTTTCAAGCCCATCATCTTTCACAGGG 240
Db 535 TTTACTGTGAAGCTCAGTGGAAACCTTCCTCTTCCTTTTCAAGCCCATCATCTTTCACAGGG 594
QY 241 GTCCTGTACAATGCCAGAGGGATTTAAAGAGGCCATGGAGCTCTTTGCTTGCAGGGTG 300
Db 595 GTCCTGTACAATGCCAGAGGGATTTAAAGAGGCCATGGAGCTCTTTGCTTGCAGGGTG 654
QY 301 CTTGGGAATTTACTCTCCAGCTTTTGATGTGTAGCTGCATCATTTGCAAGGTGAATATTTGG 360
Db 655 CTTGGGAATTTACTCTCCAGCTTTTGATGTGTAGCTGCATCATTTGCAAGGTGAATATTTGG 714
```

```
QY 361 CTAATGAGGAAGCAAAATTTTGGCTAATAAGGAAGAAATTTCTAAGCAGCAAAAGCATTTCAA 420
Db 715 CTAATGAGGAAGCAAAATTTTGGCTAATAAGGAAGAAATTTCTAAGCAGCAAAAGCATTTCAA 774
QY 421 GAGTGACTTGGTGCTGTTAAAGGCA 447
Db 775 GAGTGACTTGGTGCTGTTAAAGGCA 801

RESULT 3
US-09-997-610-3
; Sequence 3, Application US/09997610
; Patent No. US20020156244A1
; GENERAL INFORMATION:
; APPLICANT: Fox, Brian
; APPLICANT: Holloway, James L.
; TITLE OF INVENTION: ADIPOCYTE COMPLIMENT RELATED PROTEIN
; FILE REFERENCE: 00-96
; CURRENT APPLICATION NUMBER: US/09/997,610
; PRIOR FILING DATE: 2001-11-29
; PRIOR APPLICATION NUMBER: US 60/253,924
; PRIOR FILING DATE: 2000-11-29
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 3
; LENGTH: 1377
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Degenerate polynucleotide encoding a polypeptide
; OTHER INFORMATION: of SEQ ID NO:2
; NAME/KEY: variation
; LOCATION: (1)...(1377)
; OTHER INFORMATION: Each n is independently A, T, G, or C.
; NAME/KEY: misc.feature
; LOCATION: (1)...(1377)
; OTHER INFORMATION: n = A,T,C or G
US-09-997-610-3

Query Match 72.5%; Score 324.2; DB 9; Length 1377;
Best Local Similarity 61.0%; Pred. No. 1e-91;
Matches 272; Conservative 95; Mismatches 79; Indels 0; Gaps 0;

QY 1 ATAGTGGTCATACCTGCTCTTAATAACGGCAGTCATTTGAGCATGTAGAAGTTGCTGGACCT 60
Db 1 ATAGTGGTCATACCTGCTCTTNTYNATHACNGCTNATHGARGCTNGARGTNGARGTNGARGTNG 60
QY 61 CCAGCACACCCAGGCCCCAGAGAAGTGGGGCCCTCCTGGTGCCACAGGTTTACCACAA 120
Db 61 CCNGCNAYCCNMGNCNCCNGARGARGTNGNCCNCCNGNCCNCCNGNCCNCCNCCNCCNCCN 120
QY 121 TATACAGGAGAAATAAGTGAATGACAAATGCCCTGCTCCATATAGAAAGGTCAGCC 180
Db 121 TATACAGGAGAAATAAGTGAATGACAAATGCCCTGCTCCATATAGAAAGGTCAGCC 180
QY 181 TTTACTGTGAAGCTCAGTGGAAACCTTCCTCTTCCTTTTCAAGCCCATCATCTTTCACAGGG 240
Db 181 TTTACGTHAARTYNNNSNGNAAARYTNCCNTTNCNTTYAARCCNATHATHATTTACNGGN 240
QY 241 GTCCTGTACAATGCCAGAGGGATTTAAAGAGGCCATGGAGCTCTTTGCTTGCAGGGTG 300
Db 241 GTNVTNTAVAAAYCNCARMNGNAYYTNAARGARGCNATGGGNGTNTYGCNTGYMNGTN 300
QY 301 CTTGGGAATTTACTCTCCAGCTTTTGATGTGTAGCTGCATCATTTGCAAGGTGAATATTTGG 360
Db 301 CCNGNAAATYATYAYWSNNSNTTYGAYGTNGARYTNCAYCAYTGYAARGTNAAYATHTGG 360
QY 361 CTAATGAGGAAGCAAAATTTTGGCTAATAAGGAAGAAATTTCTAAGCAGCAAAAGCATTTCAA 420
Db 361 YTNATGHNARCARATHYTNCGNAAAYARGARGARATHWSNARCARCARWSNATHCAR 420
QY 421 GAGTGACTTGGTGCTGTTAAAGGC 446
```



```
QY 163 GATATAGAAAGGTGACGCTTTACTGTGAAGCTCAGTGGAAACTTCTCTCTCTTCAAG 222
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 276 TGGCCACAGACAGTCCGCCCTTCACTGTGAAGCTCAGTGGCCAGTTGCCTTCCCTTCAAAG 335

QY 223 CCATCATCTTCACAGGGTCTCTGTACATGCCACAGAGGATTTAAAGGAGGCCATGGGA 282
    || | ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 336 CCGTGTCCCTTCACAGAGTCTCTGTACATGCCACAGAGACTTTACAGGAGGACACTGGG 395

QY 283 GTCTTTGCTTGCAGGTGCC 302
    ||| ||| ||| ||| ||| |||
Db 396 GTCTTCACATGCAAGGTGCC 415

RESULT 7
US-09-960-352-220
; Sequence 220, Application US/09960352
; Patent No. US20020137139A1
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Tao, Nengbing
; APPLICANT: Byatt, John C.
; APPLICANT: Mathialagan, Nagappan
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
; FILE REFERENCE: 16511.006/37-21(10298)C
; CURRENT APPLICATION NUMBER: US/09/960,352
; CURRENT FILING DATE: 2001-09-24
; NUMBER OF SEQ ID NOS: 15112
; SEQ ID NO 220
; LENGTH: 389
; TYPE: DNA
; ORGANISM: Bos taurus
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (277)..(335)
; OTHER INFORMATION: unsure at all n locations
; OTHER INFORMATION: Clone ID: 01-LIB34-084-Q1-E1-A9
US-09-960-352-220

Query Match 20.5%; Score 91.6; DB 10; Length 389;
Best Local Similarity 76.7%; Pred. No. 4.8e-19;
Matches 112; Conservative 0; Mismatches 34; Indels 0; Gaps 0;

QY 234 CACAGGGTCTCTACAAATGCCAGAGGATTTAAAGGAGGCCATGGAGTCTTTCTGTTG 293
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1 CCCACGGTCCAGTACAAATGCCAGAGACACTTACAGGAGGACCTGGGCTTTCACATG 60

QY 294 CAGGGTCCCTGGGAATTAATCTTACTCCAGCTTTGATGTGATGTCATCATTTGCAAGTGA 353
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 61 CAGGGTGCCAGGAATTAATCACTTTCTTTGATGTGGATCTCCATCACTGCAAGGTGAC 120

QY 354 TATTTGGCTAATCAGGAAGCAATTT 379
    || | ||| ||| ||| ||| ||| |||
Db 121 TGTTCAGCTGATGAGGACAAAAGTT 146

RESULT 8
US-09-960-352-1786
; Sequence 1786, Application US/09960352
; Patent No. US20020137139A1
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Tao, Nengbing
; APPLICANT: Byatt, John C.
; APPLICANT: Mathialagan, Nagappan
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
; FILE REFERENCE: 16511.006/37-21(10298)C
; CURRENT APPLICATION NUMBER: US/09/960,352
; CURRENT FILING DATE: 2001-09-24
; NUMBER OF SEQ ID NOS: 15112
; SEQ ID NO 1786
; LENGTH: 392
```

```
; TYPE: DNA
; ORGANISM: Bos taurus
; OTHER INFORMATION: Clone ID: 08-LIB34-023-Q1-E1-B7
US-09-960-352-1786

Query Match 18.5%; Score 82.8; DB 10; Length 392;
Best Local Similarity 61.4%; Pred. No. 2.8e-16;
Matches 151; Conservative 0; Mismatches 92; Indels 3; Gaps 1;

QY 46 GAAGTTGCTGGACCTCCACACACACCCCGCCCGCCAGAGAGAGTGGGGCTCTCTGTGTGCA 105
    || | ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 146 GGACCTCCGGGGCCACACAGGACACCCAGGCGCTTCCAGGTATAGAGGGCTCCAGGTATA 205

QY 106 CCAGGTTTACCACAAATATACAGGAG---AAATAAGTGAATGACAAAATGCCCTCTCTCT 162
    || | ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 206 AGAGGAATACAGGTTTGCAGGTCCGCGCGGAATCCAGGACCAAGTGTAAATGCCCA 265

QY 163 GATATAGAAAGGTGACGCTTTTACTGTGAAGCTCAGTGGAAACTTCTCTCTTCAAG 222
    || | ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 266 TGCACACAGACAGTCCGCCCTTCACTGTGAAGCTCAGTGGCCAGTTGCCCTTCCAAAG 325

QY 223 CCATCATCTTCACAGGGTCTCTGTACATGCCACAGAGGATTTAAAGGAGGCCATGGGA 282
    || | ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 326 CCGTGTCCCTTCACAGAGTCTCTGTACAAATGCCAGAGACTTTACAGGATGACACTGGG 385

QY 283 GTCTTT 288
    || |||
Db 386 GTCTTT 391

RESULT 9
US-09-960-352-3715
; Sequence 3715, Application US/09960352
; Patent No. US20020137139A1
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Tao, Nengbing
; APPLICANT: Byatt, John C.
; APPLICANT: Mathialagan, Nagappan
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
; FILE REFERENCE: 16511.006/37-21(10298)C
; CURRENT APPLICATION NUMBER: US/09/960,352
; CURRENT FILING DATE: 2001-09-24
; NUMBER OF SEQ ID NOS: 15112
; SEQ ID NO 3715
; LENGTH: 410
; TYPE: DNA
; ORGANISM: Bos taurus
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (378)
; OTHER INFORMATION: unsure at all n locations
; OTHER INFORMATION: Clone ID: 16-LIB34-049-Q1-E1-D11
US-09-960-352-3715

Query Match 17.9%; Score 80.2; DB 10; Length 410;
Best Local Similarity 54.6%; Pred. No. 1.9e-15;
Matches 185; Conservative 0; Mismatches 148; Indels 6; Gaps 1;

QY 40 CATGTAGAAGTTGCTGGACCTCCAGCACACCCCGCCAGGAGGAGTGGGGCTCTCT 99
    || | ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 5 CCTCCAGGCTTCCAGGCGCTTCCCTGGCCCTCTCTGGCCCTCTCCAGGCGCCCA 64

QY 100 GGTGCACCGAGTTTACCAATATACAGGAGAAATAGTGAATGACAAAATGCCCTGT 159
    || | ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 65 GGCCCTCCAGGATTTGGAGGACCACTGGGATTTCCGGGAGACATTTGAGATTTGCTGTCT 124

QY 160 CCGTATATAGAAAGGTGACGCTTTTACTGTGAAGCTCAGTGGAAACTTCTCTCTCTTTC 219
    || | ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 125 CCCCCTAAA-----TCTGCTTTTGGCAGTGAAGATGAATGATCCCTCCAGGCGCCCTCC 178

QY 220 AAGCCCATCATCTTTCACAGGGGTCTGTACAATGCCAGAGGATTTAAAGGAGGCCATG 279
```


Db 438 GGAGTGTTCACCTGCACCATCCCTGG 463

RESULT 13

```
US-09-864-761-25080
; Sequence 25080, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aeonica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 25080
; LENGTH: 273
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AL121767.3
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 2
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 5.7
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 6.1
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 2
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.8
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 5.5
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 2.5
; OTHER INFORMATION: SWISSPROT HIT: P29539, EVALUE 6.80e-01
; OTHER INFORMATION: EST HUMAN HIT: BE156318.1, EVALUE 3.00e-78
; OTHER INFORMATION: NT HIT: AL163210.2, EVALUE 7.00e-75
US-09-864-761-25080
```

```
Query Match 13.1%; Score 58.4; DB 10; Length 273;
Best Local Similarity 74.0%; Pred. No. 1e-08;
Matches 74; Conservative 0; Mismatches 26; Indels 0; Gaps 0;

QY 348 GGTGAATATTGGCTAATGAGGAGCAAAATTTGGGCTTAATAAGGAAGAAATTTCTTAAGCA 407
Db 115 GGGCAACATTTGAACCTTGAGAGAGATGATTTGGGGTATCTGGTGGAGAAATTTCTTAACA 174
QY 408 GCAAAGCATTTCAAGAGGTGACTTGGGTGCTGTTAAAGGCA 447
Db 175 GCAAAGCATTTCAAGAGGTGACTTGGGTGCTGTTAAAGGCA 214
```

RESULT 14

```
US-09-864-761-8349
; Sequence 8349, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aeonica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 8349
; LENGTH: 519
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AL121767.3
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 2
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 5.7
```



```

> TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
> TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
> FILE REFERENCE: Aecomica-X-1
> CURRENT APPLICATION NUMBER: US/09/864,761
> CURRENT FILING DATE: 2001-05-23
> PRIOR APPLICATION NUMBER: US 60/180,312
> PRIOR FILING DATE: 2000-02-04
> PRIOR APPLICATION NUMBER: US 60/207,456
> PRIOR FILING DATE: 2000-05-26
> PRIOR APPLICATION NUMBER: US 09/632,366
> PRIOR FILING DATE: 2000-08-03
> PRIOR APPLICATION NUMBER: GB 24263.6
> PRIOR FILING DATE: 2000-10-04
> PRIOR APPLICATION NUMBER: US 60/236,359
> PRIOR FILING DATE: 2000-09-27
> PRIOR APPLICATION NUMBER: PCT/US01/00666
> PRIOR FILING DATE: 2001-01-30
> PRIOR APPLICATION NUMBER: PCT/US01/00667
> PRIOR FILING DATE: 2001-01-30
> PRIOR APPLICATION NUMBER: PCT/US01/00664
> PRIOR FILING DATE: 2001-01-30
> PRIOR APPLICATION NUMBER: PCT/US01/00669
> PRIOR FILING DATE: 2001-01-30
> PRIOR APPLICATION NUMBER: PCT/US01/00665
> PRIOR FILING DATE: 2001-01-30
> PRIOR APPLICATION NUMBER: PCT/US01/00668
> PRIOR FILING DATE: 2001-01-30
> PRIOR APPLICATION NUMBER: PCT/US01/00663
> PRIOR FILING DATE: 2001-01-30
> PRIOR APPLICATION NUMBER: PCT/US01/00662
> PRIOR FILING DATE: 2001-01-30
> PRIOR APPLICATION NUMBER: PCT/US01/00661
> PRIOR FILING DATE: 2001-01-30
> PRIOR APPLICATION NUMBER: PCT/US01/00670
> PRIOR FILING DATE: 2001-01-30
> PRIOR APPLICATION NUMBER: US 60/234,687
> PRIOR FILING DATE: 2000-09-21
> PRIOR APPLICATION NUMBER: US 09/608,408
> PRIOR FILING DATE: 2000-06-30
> PRIOR APPLICATION NUMBER: US 09/774,203
> PRIOR FILING DATE: 2001-01-29
> NUMBER OF SEQ ID NOS: 49117
> SOFTWARE: Annonmax Sequence Listing Engine vers. 1.1
> SEQ ID NO 2925

```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 19, 2003, 16:09:30 ; Search time 29.509 Seconds
(without alignments)
4645.518 Million cell updates/sec

Title: US-09-997-610-1_COPY_2_448
Perfect score: 447
Sequence: 1 atagtggtcataoctgtctt.....cttggtgctgttaaggca 447

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

- Database :
- 1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
 - 2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
 - 3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
 - 4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
 - 5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
 - 6: /cgn2_6/ptodata/2/ina/backfiles.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	48.4	10.8	1313	2	US-08-463-911-6
2	48.4	10.8	1517	4	US-09-140-804-9
3	42.4	9.5	1107	3	US-09-188-930-217
4	40	8.9	1276	2	US-08-463-911-1
5	37.4	8.4	728	4	US-09-336-536-2
6	37.4	8.4	1338	4	US-09-336-536-1
7	37.4	8.4	1347	4	US-09-140-804-1
8	36.2	8.1	729	4	US-09-140-804-10
9	35.8	8.0	1074	2	US-08-627-151A-15
10	35.8	8.0	1404	6	5171840-8
11	35.8	8.0	1404	6	5480796-8
12	35.8	8.0	1486	4	US-08-795-473B-3
13	35.8	8.0	1486	4	US-09-439-856-3
14	35.8	8.0	2061	6	5171840-1
15	35.8	8.0	2061	6	5480796-1
16	35.8	8.0	3319	4	US-08-795-473B-2
17	35.8	8.0	3319	4	US-09-439-856-2
18	34.4	7.7	14855	2	US-08-687-080-59
19	34.4	7.7	44453	4	US-09-146-053-5
20	34.4	7.7	49136	4	US-09-422-869-1
21	34	7.6	459	4	US-09-415-551-2
22	34	7.6	486	4	US-09-415-551-1
23	33.2	7.4	1881	4	US-09-029-348-20
24	33.2	7.4	168575	4	US-09-426-290-1
25	31.4	7.0	1001	3	US-09-188-930-218
26	31.4	7.0	1015	3	US-09-188-930-30
c 27	31.4	7.0	1395	4	US-09-149-476-88

c 28	31.4	7.0	9388	4	US-08-991-789A-141	Sequence 141, App
c 29	31.4	7.0	9388	4	US-09-062-451-141	Sequence 141, App
c 30	31.4	7.0	9388	4	US-09-598-326-141	Sequence 141, App
c 31	31.4	7.0	35081	2	US-08-752-760A-1	Sequence 1, Appl
c 32	31	6.9	1215	2	US-09-092-770-8	Sequence 8, Appl
c 33	31	6.9	1215	4	US-09-222-851-8	Sequence 8, Appl
c 34	30.8	6.9	3294	4	US-08-923-992A-7	Sequence 7, Appl
c 35	30.6	6.8	2712	1	US-08-346-455B-37	Sequence 37, Appl
c 36	30.6	6.8	2712	3	US-08-977-221-37	Sequence 37, Appl
c 37	30.6	6.8	2712	4	US-09-483-831B-37	Sequence 37, Appl
c 38	30.6	6.8	2712	5	PCT-US95-06613-37	Sequence 37, Appl
c 39	30.6	6.8	2946	1	US-08-346-455B-35	Sequence 35, Appl
c 40	30.6	6.8	2946	3	US-08-977-221-35	Sequence 35, Appl
c 41	30.6	6.8	2946	4	US-09-483-831B-35	Sequence 35, Appl
c 42	30.6	6.8	2946	5	PCT-US95-06613-35	Sequence 35, Appl
c 43	30.6	6.8	3251	1	US-08-346-455B-68	Sequence 68, Appl
c 44	30.6	6.8	3251	3	US-08-977-221-68	Sequence 68, Appl
c 45	30.6	6.8	3251	4	US-09-483-831B-68	Sequence 68, Appl

ALIGNMENTS

RESULT 1
US-08-463-911-6
; Sequence 6, Application US/08463911
; Patent No. 5869330
; GENERAL INFORMATION:
; APPLICANT: Scherer, Philipp E.
; APPLICANT: Lodish, Harvey F.
; TITLE OF INVENTION: A NOVEL SERUM PROTEIN PRODUCED
; TITLE OF INVENTION: EXCLUSIVELY IN ADIPOCYTES
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
; STREET: Two Militia Drive
; CITY: Lexington
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02173
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/463,911
; FILING DATE:
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Granahan, Patricia
; REGISTRATION NUMBER: 32,227
; REFERENCE/DOCKET NUMBER: WHI95-05
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 861-6240
; TELEFAX: (617) 861-9540
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1313 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 73..804
US-08-463-911-6

Query Match 10.8%; Score 48.4; DB 2; Length 1313;
Best Local Similarity 48.8%; Pred. No. 7.4e-06;
Matches 164; Conservative 0; Mismatches 166; Indels 6; Gaps 1;
QY 41 ATGTAGAGTTCTGACCTCCAGCACACCCAGGCCCCCAGAGAGTGGGGCCTCTG 100

Db 281 ATCCAGGTCTTATGTGCTTAAGGAGACATCGTGAAACCGGAGTACCCGGGGCTGAAG 340
QY 101 GTGCACAGGTTTACACAAATATACAGAGAAATAGTGAATGACAAAATGCCCTGTC 160
Db 341 GTCCCCGAGGCTTCCGGGAATCCAGAGGAGGAAACCTGGAGAGGTGCCT--- 397
QY 161 CTGATATAGAAAGGTGAGCCTTTTACTGTGAAGCTCAGTGGAAACCTTCCTCTTCCTTTCA 220
Db 398 ---ATGTATACCGCTCAGCATTCAGTGTGGATTGGAGACTTACGTTACTATCCCAACA 454
QY 221 AGCCATCATCTTCACAGGGTCTCTGATAATGCCAGAGGATTTAAAGGAGGCCATGG 280
Db 455 TGCCCATTCGCTTTACCAAGATCTTACAATCAGCAAAACCACTATGATGGCTCCACTG 514
QY 281 GAGTCCTTTGCTGAGGCTGCGTGGGAATTAATCTACTCCAGCTTTGATGTTGAGCTGCATC 340
Db 515 GTAAATTCACCTGCAACATTCCTGGGCTGTACTACTTTGCCCTACCACATCACAGTCTATA 574
QY 341 ATTGCAAGGTGAATATTTGGCTAATGAGGAAGCAA 376
Db 575 TGAAGGATGTAAGGTGACGCTCTTCAAGAGGACA 610

RESULT 2
US-09-140-804-9
; Sequence 9, Application US/09140804
; Patent No. 6157930
; GENERAL INFORMATION:
; APPLICANT: Sheppard, Paul O.
; APPLICANT: Humes, Jacqueline M.
; TITLE OF INVENTION: ADIPOCYTE-SPECIFIC PROTEIN HOMOLOGS
; FILE REFERENCE: 97-49
; CURRENT APPLICATION NUMBER: US/09/140,804
; CURRENT FILING DATE: 1998-08-26
; EARLIER APPLICATION NUMBER: 60/056,983
; EARLIER FILING DATE: 1997-08-26
; NUMBER OF SEQ ID NOS: 47
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 9
; LENGTH: 4517
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-140-804-9

Query Match 10.8%; Score 48.4; DB 4; Length 4517;
Best Local Similarity 48.8%; Pred. No. 1.4e-05;
Matches 164; Conservative 0; Mismatches 166; Indels 6; Gaps 1;
QY 41 ATGTAGAGTTGTGACCTCCAGCACACCCAGGCCCCAGAAAGTGGGGCTCTCTG 100
Db 235 ATCCAGGTCTTATGTCTTAAGGAGACATCGGTGAACCGGAGTACCCGGGGCTGAAG 294
QY 101 GTGCACAGGTTTACACAAATATACAGAGAAATAGTGAATGACAAAATGCCCTGTC 160
Db 295 GTCCCCGAGGCTTCCGGGAATCCAGAGGAGGAAACCTGGAGAAAGGTGCCT--- 351
QY 161 CTGATATAGAAAGGTGAGCCTTTTACTGTGAAGCTCAGTGGAAACCTTCCTCTTCCTTTCA 220
Db 352 ---ATGTATACCGCTCAGCATTCAGTGTGGGATTGGAGACTTACGTTACTATCCCAACA 408
QY 221 AGCCATCATCTTCACAGGGTCTCTGATAATGCCAGAGGATTTAAAGGAGGCCATGG 280
Db 409 TGCCCATTCGCTTTACCAAGATCTTCTACAATCAGCAAAACCACTATGATGGCTCCACTG 468
QY 281 GAGTCCTTTGCTGAGGGTGCCTGGGAATTAATCTACTCCAGCTTTGATGTTGAGCTGCATC 340
Db 469 GTAAATTCACCTGCAACATTCCTGGGCTGTACTACTTTGCCCTACCACATCACAGTCTATA 528
QY 341 ATTGCAAGGTGAATATTTGGCTAATGAGGAAGCAA 376
Db 529 TGAAGGATGTAAGGTGACGCTCTTCAAGAGGACA 564

RESULT 3
US-09-188-930-217
; Sequence 217, Application US/09188930A
; Patent No. 6150502
; GENERAL INFORMATION:
; APPLICANT: Watson, James D.
; APPLICANT: Strachan, Lorna
; APPLICANT: Sleeman, Matthew
; APPLICANT: Onrust, Rene
; APPLICANT: Murison, James Greg
; TITLE OF INVENTION: Compositions Isolated From Skin Cells
; TITLE OF INVENTION: and Methods For Their Use
; FILE REFERENCE: 11000.1011c1
; CURRENT APPLICATION NUMBER: US/09/188,930A
; CURRENT FILING DATE: 1998-11-09
; NUMBER OF SEQ ID NOS: 348
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 217
; LENGTH: 1107
; TYPE: DNA
; ORGANISM: Rat
US-09-188-930-217

Query Match 9.5%; Score 42.4; DB 3; Length 1107;
Best Local Similarity 50.5%; Pred. No. 0.00058;
Matches 103; Conservative 0; Mismatches 101; Indels 0; Gaps 0;
QY 168 AGAAGGTGAGCCTTTACTGTGAAGCTCAGTGGAAACCTTCCTCTTCCTTTCAAGCCCAT 227
Db 614 AGCCAAAGTGGCCTTTTCGGTGGCGGTAAACCAAGAGTTACCCACGTGAGCGACTGCCCAT 673
QY 228 CATCTTCACAGGGTCTCTGTACAATGCCAGAGGGATTTAAAGGAGGCCATGGAGTCTT 287
Db 674 CAAAGTTGACAGATTTCTGATGAATGAGGAGGCCACTACAATGCATGCCAGTGGCAAGTT 733
QY 288 TGCTTCAGGGTGGCTGGGAATTTACTACTCCAGCTTTGATGTTGAGCTGCATCATTTGCAA 347
Db 734 CGTCTGACAGCTGCCAGGATCTATTACTTACCTATGACATTACGCTGCGCAACACA 793
QY 348 GGTGAATATTTGGCTAATGAGGAA 371
Db 794 CCTGGCCATCGGCTAGTGCAAA 817

RESULT 4
US-08-463-911-1
; Sequence 1, Application US/08463911
; Patent No. 5869330
; GENERAL INFORMATION:
; APPLICANT: Scherer, Philipp E.
; APPLICANT: Lodish, Harvey F.
; TITLE OF INVENTION: A NOVEL SERUM PROTEIN PRODUCED
; TITLE OF INVENTION: EXCLUSIVELY IN ADIPOCYTES
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
; STREET: Two Militia Drive
; CITY: Lexington
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02173
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/463,911
; FILING DATE:
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Granahan, Patricia

Query Match 8.0%; Score 35.8; DB 4; Length 1486;
Best Local Similarity 54.1%; Pred. No. 0.091;
Matches 73; Conservative 0; Mismatches 62; Indels 0

[illegible]

```

RESULT 14
5171840-1
; Patent NO. 5171840
; APPLICANT: KISHIMOTO, TADAMITSU
; TITLE OF INVENTION: RECEPTOR PROTEIN FOR HUMAN B CELL
; STIMULATORY FACTOR-2
; NUMBER OF SEQUENCES: 11
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/298,694
; FILING DATE: 19-JAN-1989
; SEQ ID NO:1:
; LENGTH: 2061
5171840-1

```

[illegible]

RESULT 15
 5480796-1
 ; Patent No. 5480796
 ; APPLICANT: KISHIMOTO, TADAMITSU
 ; TITLE OF INVENTION: ANTIBODIES AGAINST THE RECEPTOR PROTEIN
 ; . FOR HUMAN B CELL STIMULATORY FACTOR-2
 ; NUMBER OF SEQUENCES: 8
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/07/907,650
 ; FILING DATE: 02-JUL-1992
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 298,694
 ; FILING DATE: 19-JAN-1989
 ; SEQ ID NO:1:
 ; LENGTH: 2061
 5480796-1

	Query Match	8.0%	Score 35.8;	DB 6;	Length 2061;
	Best Local Similarity	54.1%	Pred. No. 0.11;		
	Matches 73;	Conservative	0;	Mismatches 62;	Indels 0; Gaps 0;
QY	10	ATACCTGCTCTAATAACGGCGCATTCATTGAGCATCATAGAAATGCTGGACCTCCAGCACAC	69		
Db	1014	ATATCGGGCTGAACGGTCAAGAAGATCATCAAAACATGGATGGTCAAGGACCTCCAGCATCA	1073		

Qy	70	CCAGGCCCCCAAGAAAGTGGGGCCCTCCTGGTGACACAGGTTTACCAATATACAGGA	129
Db	1074	CTGTGTATCCACGACGCCCTGGAGCGCCCTGAGGCAGCTGGTGCACTTCGTGCCCAGGA	1133
Qy	130	GAATAAGTGAATG	144
Db	1134	GGAGTTCGGGCAAGG	1148

Search completed: February 19, 2003, 23:02:37
Job time : 41.509 secs

•

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 19, 2003, 15:08:07 ; Search time 154.704 Seconds
(without alignments)
6506.903 Million cell updates/sec

Title: us-09-997-610-1_copy_2_448

Perfect score: 447

Sequence: 1 atagtggtacacgtgtctt.....cttggtgtgttaaggca 447

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N_Geneseq_101002.*

- 1: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*
- 2: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*
- 3: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*
- 4: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*
- 5: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*
- 6: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*
- 7: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*
- 8: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*
- 9: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*
- 10: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*
- 11: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
- 12: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*
- 13: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*
- 14: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*
- 15: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*
- 16: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*
- 17: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*
- 18: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*
- 19: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*
- 20: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*
- 21: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*
- 22: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.*
- 23: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.*
- 24: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	351	78.5	1338	24	Human genset metab
2	58.6	13.1	404	21	Human secreted pro
3	58.4	13.1	273	22	Human foetal liver
4	58.4	13.1	273	22	Probe #18226 for g
5	58.4	13.1	273	22	Human brain expres
6	58.4	13.1	273	22	Human bone marrow
7	58.4	13.1	273	22	Probe #16759 for g
8	58.4	13.1	273	22	Probe #24296 used
9	58.4	13.1	519	22	Human foetal liver

10	58.4	13.1	519	22	ABA29883	Probe #8349 for ge
11	58.4	13.1	519	22	AAK10911	Human brain expres
12	58.4	13.1	519	22	AAK36769	Human bone marrow
13	58.4	13.1	519	22	AAI17621	Probe #7554 for ge
14	58.4	13.1	519	22	AAI42541	Probe #11227 used
15	57	12.8	662	22	AAI15063	Human breast cance
16	57	12.8	686	22	AAI23916	Human breast cance
17	57	12.8	930	23	AAI23916	DNA encoding novel
18	57	12.8	999	23	AAI23916	DNA encoding novel
19	57	12.8	1077	23	AAI23916	DNA encoding novel
20	57	12.8	1946	22	ABM44227	Human breast cell
21	57	12.8	1946	22	ABM44227	Human foetal liver
22	57	12.8	1946	22	ABA24459	Probe #2925 for ge
23	57	12.8	1946	22	AAK02964	Human brain expres
24	57	12.8	1946	22	AAK38410	Human bone marrow
25	57	12.8	1946	22	AAI12975	Probe #2908 for ge
26	57	12.8	1946	22	AAI34334	Probe #3020 used t
27	57	12.8	1946	22	AAI02893	Probe #2884 used t
28	57	12.8	1946	24	ABS02918	Human genome-deriv
C 29	56.4	12.6	32127	22	AAI99255	Human excretory re
C 30	56.4	12.6	32127	22	AAI63605	Human kidney relat
C 31	56.4	12.6	72215	22	AAK86832	Human immune/hagma
32	56	12.5	505	22	ABA60370	Human foetal liver
33	56	12.5	505	22	ABA28609	Probe #7075 for ge
34	56	12.5	505	22	AAK08648	Human brain expres
35	56	12.5	505	22	AAK34535	Human bone marrow
36	56	12.5	505	22	AAI40252	Probe #8938 used t
37	56	12.5	505	24	ABS09176	Human genome-deriv
38	56	12.5	1054	23	ABS64468	DNA encoding novel
C 39	55.4	12.4	342	21	AAI26950	Human secreted pro
40	55.4	12.4	520	22	ABA62906	Human foetal liver
41	55.4	12.4	520	22	ABA30177	Probe #8643 for ge
42	55.4	12.4	520	22	AAK11311	Human brain expres
43	55.4	12.4	520	22	AAK37101	Human bone marrow
44	55.4	12.4	520	22	AAI17942	Probe #7875 for ge
45	55.4	12.4	520	22	AAI42928	Probe #11614 used

ALIGNMENTS

RESULT 1		
AAL44066	ID	AAL44066 standard: cDNA: 1338 BP.
XX	AC	AAL44066;
XX	DT	27-SEP-2002 (first entry)
XX	DE	Human genset metabolic gene (GMG-9) cDNA sequence.
XX	DE	Human; gene: ss; gene therapy; genset metabolic gene; GMG-7B;
KW	KW	GMG-8; GMG-9; GMG-10; GMG-11; metabolic-related disorder; obesity;
KW	KW	impaired glucose tolerance; insulin resistance; Syndrome X;
KW	KW	Type II diabetes; hyperlipidaemia; atherosclerosis; hypertension;
KW	KW	heart disease; cardiac insufficiency; coronary insufficiency;
KW	KW	high blood pressure; insulin sensitivity;
KW	KW	non-insulin dependent diabetes mellitus.
OS	OS	Homo sapiens.
XX	XX	
XX	Key	Location/Qualifiers
FT	FT	1..1338
FT	FT	/tag= a
FT	FT	/partial
FT	FT	/product= "Human GMG-9 protein"
FT	FT	/note= "No stop codon is given"
XX	XX	
PN	PN	WO200255694-A2.
XX	XX	
PD	PD	18-JUL-2002.
XX	XX	
XX	XX	15-JAN-2002; 2002WO-IB01215.
PF	PF	

XX	Penn SG, Hanzel DK, Chen W, Rank DR;
XX	WPI: 2001-488899/53.
XX	Single exon nucleic acid probes for analyzing gene expression in human
PT	hearts -
XX	Claim 4: SEQ ID No 18226; 530pp; English.
XX	The present invention relates to single exon nucleic acid probes for
CC	measuring human gene expression in a sample derived from human heart. The
CC	present sequence is one such probe. The probes may be used for
CC	predicting, measuring and displaying gene expression in samples derived
CC	from the human heart via microarrays. By measuring gene expression, the
CC	probes are useful for predicting, diagnosing, grading, staging,
CC	monitoring and prognosing diseases of the human heart and vascular system
CC	e.g. cardiovascular disease, hypertension, cardiac arrhythmias and
CC	congenital heart disease.
CC	Note: The sequence data for this patent did not form part of the printed
CC	specification, but was obtained in electronic format directly from WIPO
CC	at ftp.wipo.int/pub/published_pct_sequences.
XX	
XX	Sequence 273 BP; 80 A; 36 C; 81 G; 76 T; 0 other;
XX	
QY	Query Match 13.1%; Score 58.4; DB 22; Length 273;
DB	Best Local Similarity 74.0%; Pred. No. 1.2e-08;
XX	Matches 74; Conservative 0; Mismatches 26; Indels 0; Gaps 0
QY	348 GGTGAATATTGGCTAATGAGGAACCAATTTTGGCTAATAGGAAGAAATTTCTAAGCA 407
DB	115 GGGGAACATTGAACCTGAGAGAGATGATTTGGGTATCTGGTGAAGAAATTTCTAACA 174
QY	408 GCAAGCATTCACAGAGGTGACTTGGTGCTGTAAAGGCA 447
DB	175 GCAAGCATTCACAGAGGTGACTTGGTGCTGTAAAGGCA 214
XX	
XX	RESULT 5
XX	AAK23591
ID	AAK23591 standard; DNA; 273 BP.
XX	
AC	AAK23591;
XX	
DT	05-NOV-2001 (first entry)
XX	
DE	Human brain expressed single exon probe SEQ ID NO: 23582.
XX	
KW	Human; brain expressed exon; gene expression analysis; probe;
KW	microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;
KW	epilepsy; cancer; ss.
XX	
OS	Homo sapiens.
XX	
PN	WO200157275-A2.
XX	
PD	09-AUG-2001.
XX	
XX	30-JAN-2001; 2001WO-US00667.
XX	
PR	04-FEB-2000; 2000US-0180312.
PR	26-MAY-2000; 2000US-0207456.
PR	30-JUN-2000; 2000US-0608408.
PR	03-AUG-2000; 2000US-0632366.
PR	21-SEP-2000; 2000US-0234687.
PR	27-SEP-2000; 2000US-0236359.
PR	04-OCT-2000; 2000GB-0024263.
XX	
PA	(MOLE-) MOLECULAR DYNAMICS INC.
XX	
PI	Penn SG, Hanzel DK, Chen W, Rank DR;
XX	WPI: 2001-483446/52.
XX	

XX Single exon nucleic acid probes for analyzing gene expression in human
PT brains -
XX
XX Example 4; SEQ ID NO: 23582; 650pp + Sequence Listing; English.
XX
CC The present invention provides a number of single exon nucleic acid
CC probes which are derived from genomic sequences expressed in the human
CC brain. They can be used to measure gene expression in brain cell samples,
CC which may enable the diagnosis and improved treatment of nervous system
CC diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,
CC epilepsy and cancers. The present sequence is one of the probes of the
CC invention.
XX
XX Sequence 273 BP; 80 A; 36 C; 81 G; 76 T; 0 other;
SQ
Query Match 13.1%; Score 58.4; DB 22; Length 273;
Best Local Similarity 74.0%; Pred. No. 1.2e-08;
Matches 74; Conservative 0; Mismatches 26; Indels 0; Gaps 0;
QY 348 GGTGAATATTGGCTAATGAGGAAGCAAAATTTTGGCTAATAAGGAAGAAATTTCTAAGCA 407
DB 115 GGGACATTTGAACTTGAGAGAGATGATTTGGGGTATCTGGTGAAGAAATTTCTAACA 174
QY 408 GCAAGCATTTCAAGAGGTGACTTGGGTGCTGTTAAAGGCA 447
DB 175 GCAAGCATTTCAAGAGGTGACTTGGGTGCTGTTAAAGGCA 214
RESULT 6
AAK49719
ID AAK49719 standard; DNA; 273 BP.
XX
XX AAK49719;
XX
XX 06-NOV-2001 (first entry)
DT
DE Human bone marrow expressed single exon probe SEQ ID NO: 24276.
XX
XX Human; bone marrow expressed exon; gene expression analysis; probe;
KW microarray; cancer; leukaemia; lymphoma; myeloma; ss.
XX
XX Homo sapiens.
OS
XX WO200157276-A2.
PN
XX 09-AUG-2001.
PD
XX
XX 30-JAN-2001; 2001WO-US00668.
PF
XX 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
PA
XX Penn SG, Hanzel DK, Chen W, Rank DR;
PI
XX WPI; 2001-488901/53.
DR
XX Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human bone marrow -
PT
XX Example 4; SEQ ID NO: 24276; 658pp + Sequence Listing; English.
PS
XX The present invention provides a number of single exon nucleic acid
CC probes which are derived from genomic sequences expressed in the human
CC bone marrow. They can be used to measure gene expression in bone marrow
CC samples, which may enable the improved diagnosis and treatment of cancers

CC such as lymphoma, leukaemia and myeloma. The present sequence is one of
CC the probes of the invention.
XX
XX Sequence 273 BP; 80 A; 36 C; 81 G; 76 T; 0 other;
SQ
Query Match 13.1%; Score 58.4; DB 22; Length 273;
Best Local Similarity 74.0%; Pred. No. 1.2e-08;
Matches 74; Conservative 0; Mismatches 26; Indels 0; Gaps 0;
QY 348 GGTGAATATTGGCTAATGAGGAAGCAAAATTTTGGCTAATAAGGAAGAAATTTCTAAGCA 407
DB 115 GGGACATTTGAACTTGAGAGAGATGATTTGGGGTATCTGGTGAAGAAATTTCTAACA 174
QY 408 GCAAGCATTTCAAGAGGTGACTTGGGTGCTGTTAAAGGCA 447
DB 175 GCAAGCATTTCAAGAGGTGACTTGGGTGCTGTTAAAGGCA 214
RESULT 7
AAI26826
ID AAI26826 standard; DNA; 273 BP.
XX
XX AAI26826;
AC
XX 12-OCT-2001 (first entry)
DT
DE Probe #16759 for gene expression analysis in human cervical cell sample.
XX
XX Probe; human; microarray; gene expression; cervical epithelial cell;
KW cervical cancer; ss.
XX
XX Homo sapiens.
OS
XX WO200157278-A2.
PN
XX 09-AUG-2001.
PD
XX
XX 30-JAN-2001; 2001WO-US00670.
PF
XX 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
PA
XX Penn SG, Hanzel DK, Chen W, Rank DR;
PI
XX WPI; 2001-488901/53.
DR
XX Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human cervical epithelial cells -
PT
XX Claim 25; SEQ ID NO 16759; 487pp; English.
PS
XX The present invention relates to human single exon nucleic acid probes
CC (SENP). The present sequence is one such probe. The SENPs are derived
CC from human HeLa cells. The SENPs can be used to produce a single exon
CC microarray, which can be used for measuring human gene expression in a
CC sample derived from human cervical epithelial cells. By measuring gene
CC expression, the probes are therefore useful in grading and/or staging
CC of diseases of the cervix, notably cervical cancer.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 273 BP; 80 A; 36 C; 81 G; 76 T; 0 other;
SQ
Query Match 13.1%; Score 58.4; DB 22; Length 273;
Best Local Similarity 74.0%; Pred. No. 1.2e-08;

RESULT 9

```
KW Human; gene expression; heart; microarray; vascular system; probe;
KW cardiovascular disease; hypertension; cardiac arrhythmia;
XX congenital heart disease; ss.
OS Homo sapiens.
XX
PN WO200157274-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00666.
XX
PR 04-FEB-2000; 2000US-0180312.
XX
PR 26-MAY-2000; 2000US-0207456.
XX
PR 30-JUN-2000; 2000US-0608408.
XX
PR 03-AUG-2000; 2000US-0632366.
XX
PR 21-SEP-2000; 2000US-0234687.
XX
PR 27-SEP-2000; 2000US-0236359.
XX
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX WPI; 2001-488899/53.
XX
XX Single exon nucleic acid probes for analyzing gene expression in human
XX hearts .
XX
PS Claim 1; SEQ ID No 8349; 530pp; English.
XX
CC The present invention relates to single exon nucleic acid probes for
CC measuring human gene expression in a sample derived from human heart. The
CC present sequence is one such probe. The probes may be used for
CC predicting, measuring and displaying gene expression in samples derived
CC from the human heart via microarrays. By measuring gene expression, the
CC probes are useful for predicting, diagnosing, grading, staging,
CC monitoring and prognosing diseases of the human heart and vascular system
CC e.g. cardiovascular disease, hypertension, cardiac arrhythmias and
CC congenital heart disease.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 519 BP; 170 A; 68 C; 148 G; 133 T; 0 other;

Query Match      13.1%; Score 58.4; DB 22; Length 519;
Best Local Similarity 74.0%; Pred. No. 1.7e-08;
Matches 74; Conservative 0; Mismatches 26; Indels 0; Gaps 0;

QY 348 GGTGAATATTGGCTAATGAGGACAAATTTTGGCTAATAGGAAGAAATTTCTAAGCA 407
DB 254 GGGGAACATTGACCTTGAGAGAGATGATTTGGGTATCTGGTGGAGAGAAATTTCTAACA 313

QY 408 GCAAGACATTCAAGAGGTGACTTGGTGTCTGTTAAAGGCA 447
DB 314 GCAAGACATTCAAGAGGTGACTTGGTGTCTGTTAAAGGCA 353

RESULT 11
AAK10911
ID AAK10911 standard; DNA; 519 BP.
XX
XX AAK10911;
XX
XX 05-NOV-2001 (first entry)
XX
XX Human brain expressed single exon probe SEQ ID NO: 10902.
XX
XX Human; brain expressed exon; gene expression analysis; probe;
KW microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;
KW epilepsy; cancer; ss.
XX
```

```
OS Homo sapiens.
XX
PN WO200157275-A2.
XX
XX 09-AUG-2001.
XX
XX 30-JAN-2001; 2001WO-US00667.
XX
XX 04-FEB-2000; 2000US-0180312.
XX
XX 26-MAY-2000; 2000US-0207456.
XX
XX 30-JUN-2000; 2000US-0608408.
XX
XX 03-AUG-2000; 2000US-0632366.
XX
XX 21-SEP-2000; 2000US-0234687.
XX
XX 27-SEP-2000; 2000US-0236359.
XX
XX 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX WPI; 2001-483446/52.
XX
XX Single exon nucleic acid probes for analyzing gene expression in human
XX brains .
XX
XX Example 4; SEQ ID NO: 10902; 650pp + Sequence Listing; English.
XX
CC The present invention provides a number of single exon nucleic acid
CC probes which are derived from genomic sequences expressed in the human
CC brain. They can be used to measure gene expression in brain cell samples,
CC which may enable the diagnosis and improved treatment of nervous system
CC diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,
CC epilepsy and cancers. The present sequence is one of the probes of the
CC invention.
XX
SQ Sequence 519 BP; 170 A; 68 C; 148 G; 133 T; 0 other;

Query Match      13.1%; Score 58.4; DB 22; Length 519;
Best Local Similarity 74.0%; Pred. No. 1.7e-08;
Matches 74; Conservative 0; Mismatches 26; Indels 0; Gaps 0;

QY 348 GGTGAATATTGGCTAATGAGGACAAATTTTGGCTAATAGGAAGAAATTTCTAAGCA 407
DB 254 GGGGAACATTGACCTTGAGAGAGATGATTTGGGTATCTGGTGGAGAGAAATTTCTAACA 313

QY 408 GCAAGACATTCAAGAGGTGACTTGGTGTCTGTTAAAGGCA 447
DB 314 GCAAGACATTCAAGAGGTGACTTGGTGTCTGTTAAAGGCA 353

RESULT 12
AAK36769
ID AAK36769 standard; DNA; 519 BP.
XX
XX AAK36769;
XX
XX 06-NOV-2001 (first entry)
XX
XX Human bone marrow expressed single exon probe SEQ ID NO: 11326.
XX
XX Human; bone marrow expressed exon; gene expression analysis; probe;
KW microarray; cancer; leukaemia; lymphoma; myeloma; ss.
XX
XX Homo sapiens.
XX
XX WO200157276-A2.
XX
XX 09-AUG-2001.
XX
XX 30-JAN-2001; 2001WO-US00668.
XX
XX 04-FEB-2000; 2000US-0180312.
XX
XX 26-MAY-2000; 2000US-0207456.
XX
```

```
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX WPI; 2001-488900/53.
XX
XX Human genome-derived single exon nucleic acid probes useful for
XX analyzing gene expression in human bone marrow -
XX
XX Example 4; SEQ ID NO: 11326; 658pp + Sequence Listing; English.
XX
XX The present invention provides a number of single exon nucleic acid
XX probes which are derived from genomic sequences expressed in the human
XX bone marrow. They can be used to measure gene expression in bone marrow
XX samples, which may enable the improved diagnosis and treatment of cancers
XX such as lymphoma, leukemia and myeloma. The present sequence is one of
XX the probes of the invention.
XX
XX Sequence 519 BP; 170 A; 68 C; 148 G; 133 T; 0 other;
XX
XX Query Match 13.1%; Score 58.4; DB 22; Length 519;
XX Best Local Similarity 74.0%; Pred. No. 1.7e-08;
XX Matches 74; Conservative 0; Mismatches 26; Indels 0; Gaps 0;
XX
Qy 348 GGTGAATATTTGGCTAATGAGGAAGCAAAATTTGGCTAATGAGGAAGAAATTTCTAAGCA 407
Db ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| ||
Qy 408 GCAAGCATTCAGAGGTCGACTTGGCTGCTGTTAAAGGCA 447
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||
Qy 314 GCAAGCATTCAGAGGTCGACTTGGCTGCTGTTAAAGGCA 353
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||

RESULT 13
AA117621
ID AA117621 standard; DNA; 519 BP.
XX
XX AC AA117621;
XX
XX 12-OCT-2001 (first entry)
XX
XX Probe #7554 for gene expression analysis in human cervical cell sample.
XX
XX Probe; human; microarray; gene expression; cervical epithelial cell;
XX cervical cancer; ss.
XX
XX Homo sapiens.
XX
XX WO200157278-A2.
XX
XX 09-AUG-2001.
XX
XX 30-JAN-2001; 2001WO-US00670.
XX
XX 04-FEB-2000; 2000US-0180312.
XX 26-MAY-2000; 2000US-0207456.
XX 30-JUN-2000; 2000US-0608408.
XX 03-AUG-2000; 2000US-0632366.
XX 21-SEP-2000; 2000US-0234687.
XX 27-SEP-2000; 2000US-0236359.
XX 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX WPI; 2001-488901/53.
XX
```

```
XX Human genome-derived single exon nucleic acid probes useful for
XX analyzing gene expression in human cervical epithelial cells -
XX
XX Claim 25; SEQ ID No 7554; 487pp; English.
XX
XX The present invention relates to human single exon nucleic acid probes
XX (SENPs). The present sequence is one such probe. The SENPs are derived
XX from human HeLa cells. The SENPs can be used to produce a single exon
XX microarray, which can be used for measuring human gene expression in a
XX sample derived from human cervical epithelial cells. By measuring gene
XX expression, the probes are therefore useful in grading and/or staging
XX of diseases of the cervix, notably cervical cancer.
XX Note: The sequence data for this patent did not form part of the printed
XX CC specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 519 BP; 170 A; 68 C; 148 G; 133 T; 0 other;
XX
XX Query Match 13.1%; Score 58.4; DB 22; Length 519;
XX Best Local Similarity 74.0%; Pred. No. 1.7e-08;
XX Matches 74; Conservative 0; Mismatches 26; Indels 0; Gaps 0;
XX
Qy 348 GGTGAATATTTGGCTAATGAGGAAGCAAAATTTGGCTAATGAGGAAGAAATTTCTAAGCA 407
Db ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| ||
Qy 408 GCAAGCATTCAGAGGTCGACTTGGCTGCTGTTAAAGGCA 447
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||
Qy 314 GCAAGCATTCAGAGGTCGACTTGGCTGCTGTTAAAGGCA 353
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||

RESULT 14
AA142541
ID AA142541 standard; DNA; 519 BP.
XX
XX AC AA142541;
XX
XX 17-OCT-2001 (first entry)
XX
XX Probe #11227 used to measure gene expression in human placenta sample.
XX
XX Probe; microarray; human; placenta; antenatal diagnosis;
XX genetic disorder; ss.
XX
XX Homo sapiens.
XX
XX WO200157272-A2.
XX
XX 09-AUG-2001.
XX
XX 30-JAN-2001; 2001WO-US00663.
XX
XX 04-FEB-2000; 2000US-0180312.
XX 26-MAY-2000; 2000US-0207456.
XX 30-JUN-2000; 2000US-0608408.
XX 03-AUG-2000; 2000US-0632366.
XX 21-SEP-2000; 2000US-0234687.
XX 27-SEP-2000; 2000US-0236359.
XX 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX WPI; 2001-488897/53.
XX
XX Human genome-derived single exon nucleic acid probes useful for
XX analyzing gene expression in human placenta -
XX
XX Claim 25; SEQ ID No 11227; 654pp; English.
XX
XX The present invention relates to single exon nucleic acid probes (SENPs).
```


GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 19, 2003, 23:16:26 ; Search time 1425.16 Seconds
(without alignments)
9128.055 Million cell updates/sec

Title: US-09-997-610-1_COPY_2_448

Perfect score: 447

Sequence: 1 atagtgatcatactgtctt.....cttggtgctgttaaaaggca 447

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl.*

- 1: gb_ba.*
- 2: gb_htg.*
- 3: gb_in.*
- 4: gb_on.*
- 5: gb_ov.*
- 6: gb_pat.*
- 7: gb_ph.*
- 8: gb_pl.*
- 9: gb_pr.*
- 10: gb_rst.*
- 11: gb_sy.*
- 12: gb_sy.*
- 13: gb_un.*
- 14: gb_vl.*
- 15: em_ba.*
- 16: em_fun.*
- 17: em_hum.*
- 18: em_in.*
- 19: em_mu.*
- 20: em_on.*
- 21: em_or.*
- 22: em_ov.*
- 23: em_pat.*
- 24: em_ph.*
- 25: em_pl.*
- 26: em_ro.*
- 27: em_sts.*
- 28: em_un.*
- 29: em_vl.*
- 30: em_htg_hum.*
- 31: em_htg_inv.*
- 32: em_htg_other.*
- 33: em_htg_mus.*
- 34: em_htg_pln.*
- 35: em_htg_rod.*
- 36: em_htg_mam.*
- 37: em_htg_vrt.*
- 38: em_sy.*
- 39: em_htgo_hum.*
- 40: em_htgo_mus.*
- 41: em_htgo_other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	351	78.5	145880	9	HS302D9	282198 Human DNA s
2	151.2	33.8	1265	10	TMSHP20A	D12974 Tamias asia
3	139.2	31.1	5121	10	AB067779	AB067779 Tamias si
4	84.6	18.9	2005	10	TMSHP25	D12975 Tamias asia
5	68.6	15.3	559	10	AB067813S3	AB069679 Callosciu
6	68.2	15.3	145880	9	HS302D9	282198 Human DNA s
7	63.4	14.2	90497	9	HS7328A	AL008722 Human DNA
8	63.4	14.2	135880	2	AC104256	AC104256 Homo sapi
9	63.2	14.1	51000	9	AP005265	AP005265 Homo sapi
10	63.2	14.1	144514	2	AC027433	AC027433 Homo sapi
11	63.2	14.1	149038	2	AC025112	AC025112 Homo sapi
12	62.6	14.0	162063	9	AC015553	AC015553 Homo sapi
13	62.6	14.0	175134	2	AL161623	AL161623 Homo sapi
14	62.6	14.0	176932	9	AL157884	AL157884 Homo sapi
15	62.4	14.0	1385	10	TMSHP27	D12976 Tamias asia
16	62	13.9	203234	9	AL441943	AL441943 Human DNA
17	61.8	13.8	69937	2	AC069566	AC069566 Homo sapi
18	61.8	13.8	122231	2	AC026099	AC026099 Homo sapi
19	61.8	13.8	143872	9	AC091987	AC091987 Homo sapi
20	61.8	13.8	172792	9	AC062031	AC062031 Homo sapi
21	61.8	13.8	179848	9	AC092967	AC092967 Homo sapi
22	61.8	13.8	180038	2	AC026870	AC026870 Homo sapi
23	61.2	13.7	177386	2	AC018971	AC018971 Homo sapi
24	60.2	13.5	108511	9	AC011418	AC011418 Homo sapi
25	60.2	13.5	137591	2	AL672086	AL672086 Homo sapi
26	60.2	13.5	158587	9	AC074325	AC074325 Homo sapi
27	60.2	13.5	158811	2	AC013503	AC013503 Homo sapi
28	60.2	13.5	160151	9	AC098595	AC098595 Homo sapi
29	60.2	13.5	161049	2	AL591842	AL591842 Homo sapi
30	60.2	13.5	173767	9	AC002449	AC002449 Human PAC
31	60.2	13.5	175765	9	AC027128	AC027128 Homo sapi
32	60.2	13.5	179484	9	AC107626	AC107626 Homo sapi
33	60.2	13.5	187201	2	AC068063	AC068063 Homo sapi
34	60.2	13.5	187640	9	AC073840	AC073840 Homo sapi
35	60.2	13.5	188622	9	AL359258	AL359258 Human DNA
36	60.2	13.5	192973	9	AC023471	AC023471 Homo sapi
37	60.2	13.5	207600	9	AC024910	AC024910 Homo sapi
38	60.2	13.5	222542	9	AC022379	AC022379 Homo sapi
39	59.8	13.4	170200	9	AC008012	AC008012 Homo sapi
40	59.6	13.3	144057	2	AC074266	AC074266 Homo sapi
41	59.6	13.3	172349	9	AC019185	AC019185 Homo sapi
42	59.6	13.3	195594	2	AC096992	AC096992 Homo sapi
43	59.6	13.3	197078	2	AC117382	AC117382 Homo sapi
44	59.6	13.3	233877	9	AC093798	AC093798 Homo sapi
45	59.4	13.3	98070	9	AC025463	AC025463 Homo sapi

ALIGNMENTS

RESULT 1
HS302D9
LOCUS
DEFINITION Human DNA sequence from clone RPL-302D9 on chromosome 22 Contains
GSSs, complete sequence.
ACCESSION 282198
VERSION 282198.2 GI:6572207
KEYWORDS HTG.
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 145880)
AUTHORS Bridgeman,A.
TITLE Direct Submission

JOURNAL

Submitted (08-DEC-1999) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk

On Dec 13, 1999 this sequence version replaced gi:3164067.

During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

This sequence has been finished according to sequence map criteria as follows. An attempt is made to resolve all sequencing problems, such as compressions and repeats, but not necessarily within known annotated human repeat sequence elements (e.g. Alu). Where the sequence is ambiguous, there is an annotation using the 'unsure' feature key.

The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em.; EMBL; Sw.; SWISSPROT; Tr.; TREMBL; Wp.; WORMPEP; Information on the WORMPEP database can be found at <http://www.sanger.ac.uk/projects/C.elegans/wormpep> This sequence was generated from part of bacterial clone contigs of human chromosome 22, constructed by the Sanger Centre Chromosome 22 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/Chr22>

RP1-302D9 is from the library RPci-1 constructed at the Roswell Park Cancer Institute by the group of Pieter de Jong. For further details see <http://bacpac.med.buffalo.edu/> VECTOR: pCYPAC2

This sequence is the entire insert of clone RP1-302D9 The true left end of clone CTA-282P2 is at 69682 in this sequence. The true right end of clone CTA-415G2 is at 55167 in this sequence.

FEATURES

Source	
1..145880 Homo sapiens /organism="Homo sapiens" /db_xref="taxon:9606" /chromosomes="22" /clone="RP1-302D9" /clone_lib="RPci-1"	
repeat_region 188..245 /note="MER3 repeat: matches 144..209 of consensus"	repeat_region 6647..6685 /note="MADE1 repeat: matches 1..23 of consensus"
repeat_region 246..571 /note="AluSx repeat: matches 1..312 of consensus"	repeat_region 686..6987 /note="AluSx repeat: matches 1..302 of consensus"
repeat_region 572..759 /note="MER3 repeat: matches 1..144 of consensus"	repeat_region 698..7036 /note="MADE1 repeat: matches 23..77 of consensus"
repeat_region 783..933 /note="MER5A repeat: matches 26..187 of consensus"	repeat_region 7482..7754 /note="AluJb repeat: matches 9..290 of consensus"
repeat_region 1033..1336 /note="AluSp repeat: matches 1..299 of consensus"	repeat_region 775..8060 /note="AluJo repeat: matches 1..295 of consensus"
repeat_region 1450..1583 /note="MIR repeat: matches 24..160 of consensus"	repeat_region 8414..8551 /note="L2 repeat: matches 2553..2706 of consensus"
repeat_region 1687..1752 /note="L2 repeat: matches 2593..2661 of consensus"	repeat_region 8914..9030 /note="MIR repeat: matches 147..262 of consensus"
repeat_region 2350..2660 /note="AluSc repeat: matches 3..309 of consensus"	repeat_region 9110..9280 /note="MIR repeat: matches 91..262 of consensus"
repeat_region 2684..2981 /note="AluSq repeat: matches 2..300 of consensus"	repeat_region 9283..9412 /note="MIR repeat: matches 15..144 of consensus"
repeat_region 3323..3343 /note="MLTLE repeat: matches 116..136 of consensus"	repeat_region 9521..9679 /note="FAM repeat: matches 3..161 of consensus"
repeat_region 3344..3652 /note="AluY repeat: matches 1..309 of consensus"	repeat_region 9820..10225 /note="MSRB repeat: matches 2..425 of consensus"
repeat_region 3653..3928 /note="MLTLE repeat: matches 136..359 of consensus"	repeat_region 10312..10383 /note="MIR repeat: matches 79..150 of consensus"
repeat_region 3929..4278 /note="THELB repeat: matches 3..364 of consensus"	repeat_region 10718..11310 /note="match: GSS: Em:B14024"
repeat_region 4279..4485 /note="MLTLE repeat: matches 359..568 of consensus"	repeat_region 10784..11201 /note="match: GSS: Em:B43656"
repeat_region 5073..5176 /note="52 copies 2 mer ct 78 conserved"	repeat_region 11838..11946 /note="MIR repeat: matches 20..137 of consensus"
repeat_region 5181..5491 /note="AluJb repeat: matches 1..311 of consensus"	repeat_region 12174..12445 /note="L2 repeat: matches 1988..2275 of consensus"
repeat_region 6369..6485 /note="L2 repeat: matches 2579..2705 of consensus"	misc_feature 13017..13369 /note="match: STS: Em:G49301"
	repeat_region 13331..13397 /note="MIR repeat: matches 174..244 of consensus"
	repeat_region 13398..13698 /note="AluSp repeat: matches 1..302 of consensus"
	repeat_region 13699..13810 /note="MIR repeat: matches 76..174 of consensus"
	repeat_region 13806..13919 /note="MIR repeat: matches 77..189 of consensus"
	repeat_region 13945..14060 /note="MIR repeat: matches 24..142 of consensus"
	repeat_region 14061..14367 /note="AluY repeat: matches 1..301 of consensus"
	repeat_region 14368..14452 /note="MIR repeat: matches 141..225 of consensus"
	misc_feature 14589..14679 /note="MIR repeat: matches 173..262 of consensus"
	misc_feature 14597..15201 /note="match: GSS: Em:AQ553482"
	misc_feature 14616..15060 /note="match: GSS: Em:AQ370601"
	repeat_region 14868..15040 /note="MIR repeat: matches 49..233 of consensus"
	repeat_region 15071..15188 /note="L2 repeat: matches 2112..2239 of consensus"
	repeat_region 15304..15399 /note="MLT1B repeat: matches 1..99 of consensus"
	repeat_region 15490..15662 /note="AluS91 repeat: matches 2..114 of consensus"
	repeat_region 15669..15727 /note="MLT1B repeat: matches 119..178 of consensus"
	repeat_region 15728..16027

```
repeat_region /note="AluSc repeat: matches 1. .299 of consensus"
16028. .16245
/note="MLTIB repeat: matches 178. .390 of consensus"
16546. .16854
/note="AluY repeat: matches 1. .300 of consensus"
18296. .18323
/note="MSTA repeat: matches 2. .29 of consensus"
18324. .18392
/note="MER66-internal repeat: matches 4919. .4993 of
consensus"
18393. .18712
/note="AluJb repeat: matches 1. .311 of consensus"
18713. .19133
/note="MER66-internal repeat: matches 4548. .4919 of
consensus"
complement(18872. .19230)
/note="match: GSS: Em:AQ005063"
19231. .19719
/note="match: GSS: Em:B14179"
19537. .20290
/note="HERVH21 repeat: matches 4657. .5784 of consensus"
20317. .20382
/note="33 copies 2 mer ta 68 conserved"
20513. .20666
/note="77 copies 2 mer tt 70 conserved"
20682. .21008
/note="AluSg1 repeat: matches 1. .306 of consensus"
21239. .21553
/note="HUES-P3 repeat: matches 4410. .4713 of consensus"
21882. .22254
/note="THEIB repeat: matches 1. .364 of consensus"
22302. .22537
/note="MER66-internal repeat: matches 2186. .2417 of
consensus"
22538. .22850
/note="AluSp repeat: matches 1. .313 of consensus"
22851. .23801
/note="MER66-internal repeat: matches 1210. .2186 of
consensus"
23905. .23989
/note="MER66-internal repeat: matches 3017. .3102 of
consensus"
Query Match 78.5%; Score 351; DB 9; Length 145880;
Best Local Similarity 98.6%; Pred. No. 3.8e-93;
Matches 354; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
Qy 89 TGGGCGCTCCTGGTCACCGAGTTTACCACATATACAGGAGAAATAGTGAATGACAA 148
Db 37254 TGGGAATGTCCTGGTCCCGAGTTTACCACATATACAGGAGAAATAGTGAATGACAA 37313
Qy 149 AATGCCCTGTCTCATATAGAAAGGTGAGCTTTACTGTGAAGCTCAGTGGAAATCTTC 208
Db 37314 AATGCCCTGTCTCATATAGAAAGGTGAGCTTTACTGTGAAGCTCAGTGGAAATCTTC 37373
Qy 209 CTCCTTCCTTCAAGCCCATCATCTTACAGGGGTCCTGTACAAATCCCAGAGGGATTAA 268
Db 37374 CTCCTTCCTTCAAGCCCATCATCTTACAGGGGTCCTGTACAAATCCCAGAGGGATTAA 37433
Qy 269 AGGAGCCATGGGAGCTCTTGGTCGAGGGTCCCTGGGAATTTACTACTCCAGCTTTGATG 328
Db 37434 AGGAGCCATGGGAGCTCTTGGTCGAGGGTCCCTGGGAATTTACTACTCCAGCTTTGATG 37493
Qy 329 TTGAGCTGCATTTGCAAGGTGAATATTTGGCTAATGAGGAACAAATTTTGGCTAATA 388
Db 37494 TTGAGCTGCATTTGCAAGGTGAATATTTGGCTAATGAGGAACAAATTTTGGCTAATA 37553
Qy 389 AGGAAGAAATTTCTAAGCAGCAAAAGCATTTCAAGAGGTGACCTTGGGTGCTGTTTAAAGGCA 447
Db 37554 AGGAAGAAATTTCTAAGCAGCAAAAGCATTTCAAGAGGTGACCTTGGGTGCTGTTTAAAGGCA 37612
RESULT 2
TMSHP20A
```

```
LOCUS TMSHP20A 1265 bp mRNA llnear ROD 03-FEB-1999
DEFINITION Tamias asiaticus mRNA for Hp-20, complete cds.
ACCESSION D12974
VERSION D12974.1 GI:287467
KEYWORDS HP-20; collagen-like domain; hibernation-related protein; plasma
protein.
SOURCE Tamias asiaticus liver cDNA to mRNA, clone:PCW20-7.
ORGANISM Tamias sibiricus
Mammalia; Eutheria; Rodentia; Sciurognathi; Scuridae; Sclurinae;
Tamias.
REFERENCE 1 (bases 1 to 1265)
AUTHORS Takamatsu,N., Ohba,K., Kondo,J., Kondo,N. and Shlba,T.
TITLE Hibernation-associated gene regulation of plasma proteins with a
collagen-like domain in mammalian hibernators
JOURNAL Mol. Cell. Biol. 13 (3), 1516-1521 (1993)
MEDLINE 93180798
REFERENCE 2 (bases 1 to 1265)
AUTHORS Takamatsu,N.
TITLE Direct Submission
JOURNAL Submitted (21-AUG-1992) Nobuhiko Takamatsu, School of Hygienic
Sciences, Kitasato University, Laboratory of Molecular Biology,
1-15-1 Kitasato, Sagamihara, Kanagawa 228, Japan (Tel:0427-78-9408,
Fax:0427-78-9403)
FEATURES
source
Location/Qualifiers
1..1265
/organism="Tamias sibiricus"
/db_xref="taxon:64680"
/clone="PCW20-7"
/tissue_type="liver"
89..679
/note="collagen-like domain at nt 161-277"
CDS
/codon_start=1
/product="Hp-20"
/protein_id="BAA02351.1"
/db_xref="GI:287468"
/translation="MTDVMRLAIFVLNVLDQVSCGPPGPGVPGVPGVPGPRGP
PGQFCAAGRPDGPKPSKPCRSATFKFSGRLPPSPPEVYFVLYNTQDL
KESTGVFNCVPGNYHFSFDELVHCKYKIGLKNHIOVMEKHQLSKNEYENASGAMI
MPLRQGDVWLEADVETEEPQAKVVIYFSGLISS"
89..157
sig_peptide
mat_peptide
158..676
/polyA_signal
1238..1243
polyA_site
1265
BASE COUNT 354 a 304 c 288 g 319 t
ORIGIN
Query Match 33.8%; Score 151.2; DB 10; Length 1265;
Best Local Similarity 67.8%; Pred. No. 7.3e-34;
Matches 229; Conservative 0; Mismatches 103; Indels 6; Gaps 1;
Qy 40 CATGTAGAAGTTGCTGGACCTCAGCACACCCAGGCCCCAGAGAAAGTGGGGCTCCT 99
Db 173 CCGTGTAGGATATCCAGGAGTGGCCGGGCTCGTGGCCCCAGGTCAACCA 232
Qy 100 GGTGACACAGGTGTACCAACAATATACAGAGAAATAGTGAATGACAAATGCCCTGT 159
Db 233 GCGCTGGGGTAGCCAGGAGACCCAGAGCCAAAGGCCAGCTGAATAATGCCCTG- 291
Qy 160 CCGTATATAGAAAGTTCAGCCTTTACTGTGAAGCTCAGTGGAAACCTTCCTCTTCTTC 219
Db 292 -----CAGAGAGAGTTCAGCCTTCACGGTGAAGTTTCAGCGGAGGCTCCCTCCACCTTCG 346
Qy 220 AAGCCCATCATCTTCACAGGGTCTCTGTAACAATGCCAGAGGATTTAAGAGGCCATG 279
Db 347 GAGCCTGTGTCTTCACAGAGTCTCTGTACAATACCCAGAGGACTTGAAGGAGGACACA 406
Qy 280 GGAGTCTTGTGTTCAGGGTCTGGGAATTTACTCCAGCTTTGATCTTGAGCTGCAT 339
Db 407 GGAGTCTTAACTCGGTGGAGCCTGGAATTTACCATTTTCAGCTTTGATCTTGAGCTTTAC 466
Qy 340 CATTGCAAGGTGAATATTTGGCTAATGAGGAAGCAAT 377
```

```

Db 467 CACTGCAAGGTGAAGATTGGTGGTGGTGAAGAACACCAT 504
|||||
Db 4141 GGAGTCTTTAACTGCGCTGGAGCCTGGAATTACATTTGAGCTTTGATGTTGAGCTTAC 4200
|||||
QY 340 CATTGCAAGGTGAATATTGGCTTAATGAGGAACCAAT 377
|||||
Db 4201 CACTGCAAGGTGAAGATTGGTGGTGGTGAAGAACACCAT 4238
|||||

RESULT 4
TMSHP25 2005 bp mRNA linear ROD 03-FEB-1999
LOCUS Tamias asiaticus mRNA for HP-25, complete cds.
DEFINITION
ACCESSION D12975
VERSION D12975.1 GI:287469
KEYWORDS HP-25; collagen-like domain; hibernation-related protein; plasma protein.
SOURCE Tamias asiaticus liver cDNA to mRNA, clone:PCM25-3.
ORGANISM Tamias sibiricus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Scuridae; Sciurinae; Tamias.
REFERENCE 1 (bases 1 to 2005)
AUTHORS Takamatsu,N., Ohba,K., Kondo,J., Kondo,N. and Shiba,T.
TITLE Hibernation-associated gene regulation of plasma proteins with a collagen-like domain in mammalian hibernators
JOURNAL Mol. Cell. Biol. 13 (3), 1516-1521 (1993)
MEDLINE 93180798
REFERENCE 2 (bases 1 to 2005)
AUTHORS Takamatsu,N.
TITLE Direct Submission
JOURNAL Submitted (21-AUG-1992) Nobuhiko Takamatsu, School of Hygienic Sciences, Kitasato University, Laboratory of Molecular Biology; 1-15-1 Kitasato, Sagamihara, Kanagawa 228, Japan (Tel:0427-78-9408, Fax:0427-78-9403)
FEATURES
source Location/Qualifiers
1..2005
/organism="Tamias sibiricus"
/db_xref="taxon:64680"
/clone="pCM25-3"
/tissue_type="liver"
114..761
/note="collagen-like domain at nt 231-356"
/codon_start=1
/product="HP-25"
/protein_id="BAA02352.1"
/db_xref="GI:287470"
/translation="MPAQRGGALSMGAGFWLVLISALSADSNQNSPCGPPGP
PQPGIPGPGAPGALGPPGPGVFGIPGQPGDVERKSRPSKSAFAVKLSERPPE
PFQPTVFFKALYNQEGHFNMATGEFSCVLPVYVNFQFDRLRFSQSVKILRMDGQVR
EKEAQNDSYKHAMGSVIMALGDKVWLESKLKGTSESEKGIHIVFFGYLLYK"
sig_peptide 114..197
mat_peptide 198..758
polyA_signal /product="HP-25"
polyA_site 1983..1988
BASE COUNT 569 a 472 c 494 g 470 t
ORIGIN
Query Match 18.9%; Score 84.6; DB 10; Length 2005;
Best Local Similarity 54.4%; Pred. No. 5.1e-14;
Matches 193; Conservative 0; Mismatches 159; Indels 3; Gaps 1;
QY 45 AGAAGTGTCTGGACCTCCAGACACCCCGAGGAGAGTGGGGCTCTCGTGC 104
|||||
Db 248 AGACCTCCAGCATCCAGGCTTCCAGCGCTCTCGAGCACTTGGTCCACGAGACC 307
|||||
QY 105 ACCAGTGTACCAATATACAGGAGAAATAGTGAATGACAAA---ATGCCCTGTGCC 161
|||||
Db 308 CCCAGTGTGCCAGGAATACCAGGACCAAGGTCCTCTCGAGAGCTTGAGAAGTGCTC 367
|||||
QY 162 TGATATAGAAAGGTGAGCCCTTTACTGTGAGCTCAGTGGAAACTTCCTCTCCTTCAA 221
|||||
Db 368 ATCGCGACCAATCTGCCTTTGCTGAGCTGAGTGAACGCCGCCAGAGCCCTTCCA 427
|||||

```

```

Db 467 CACTGCAAGGTGAAGATTGGTGGTGGTGAAGAACACCAT 504
|||||
Db 4141 GGAGTCTTTAACTGCGCTGGAGCCTGGAATTACATTTGAGCTTTGATGTTGAGCTTAC 4200
|||||
QY 340 CATTGCAAGGTGAATATTGGCTTAATGAGGAACCAAT 377
|||||
Db 4201 CACTGCAAGGTGAAGATTGGTGGTGGTGAAGAACACCAT 4238
|||||

RESULT 3
AB067779 5121 bp DNA linear ROD 26-OCT-2001
LOCUS Tamias sibiricus gene for HP-20, complete cds.
DEFINITION
ACCESSION AB067779
VERSION AB067779.1 GI:15706341
KEYWORDS
SOURCE Tamias sibiricus DNA, clone:lambdaCM20G1.
ORGANISM Tamias sibiricus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Scuridae; Sciurinae; Tamias.
REFERENCE 1
AUTHORS Ono,M., Hosoe,Y., Azuma,S., Shoji,M., Nara,K., Kondo,N., Shiba,T. and Takamatsu,N.
TITLE HNF-1 regulates the liver-specific transcription of the chipmunk HP-20 gene
JOURNAL Gene 277 (1-2), 121-127 (2001)
MEDLINE 21488336
REFERENCE 2 (bases 1 to 5121)
AUTHORS Ono,M.
TITLE Direct Submission
JOURNAL Submitted (04-AUG-2001) Motoharu Ono, Kitasato University, Department of Biosciences, School of Science; 1-15-1 Kitasato, Sagamihara, Kanagawa 228-8555, Japan
Sagamihara, Kanagawa 228-8555, Japan
(E-mail:ms9805mestu.sci.kitasato-u.ac.jp, Tel:81427789408, Fax:81427789408)
FEATURES
source Location/Qualifiers
1..5121
/organism="Tamias sibiricus"
/db_xref="taxon:64680"
/clone="lambdaCM20G1"
1518..1684
/note="synonym:Tamias asiaticus"
join(1678..1684,2573..2710,3968..4413)
/feature="HP-20"
join(1678..1684,2573..2710,3968..4413)
/feature="HP-20"
/codon_start=1
/product="HP-20"
/protein_id="BAB68362.1"
/db_xref="GI:15706342"
/translation="MTDAWRLAIFVLVNVNLDQVSCSPGPGVPGVPGVPRGP
PQPGAAAGRPDGPVKPSVKPCPCHRSATFKSGRLPPSPPEVFTVLYNTORDL
KASTGVNFCVPGNHFSDVELYHCKYKIGLKNHIOVMEKHQLSKNEYENASGAMI
MPLRQGRVWLEADVETEPDQAKVVIYFSGFLISS"
exon 2573..2710
exon /gene="HP-20"
exon 3968..5000
BASE COUNT 1571 a 1018 c 1090 g 1442 t
ORIGIN
Query Match 31.1%; Score 139.2; DB 10; Length 5121;
Best Local Similarity 71.6%; Pred. No. 3.1e-30;
Matches 199; Conservative 0; Mismatches 73; Indels 6; Gaps 1;
QY 100 GGTGCACCAAGTTTACCACAATATACAGGAGAAATAGTGAATGACAAAATGCCCTCT 159
|||||
Db 3967 GGGCGTGGCGTAGGCGAGAGACCCAGCAACCAAGGCCAGTGTAATAATGCCCTG- 4025
|||||
QY 160 CCTGATATAGAAGGTCACCCCTTTACTGTGAAGCTCAGTGGAAACTTCTCTTCCTTTC 219
|||||
Db 4026 ----CAGAGAGAGGTCACCCCTTACGGTGAAGTTCAGCGGAAGGCTCCCTCCACCTCG 4080
|||||
QY 220 AAGCCCATCATCTTACAGGGGTCTCTGTACAAATGCCAGGAGGATTTAAAGGAGGCGCATG 279
|||||
Db 4081 GAGCCTGTGCTTCTACAGAGGTCCTGTACAAATACCCAGAGGACTTGAGGCGGACACA 4140
|||||
QY 280 GGAGTCTTTGCTGAGGCGTCTGGGAATTACTACTCCAGCTTTTGATGTTGAGCTGCAAT 339
|||||

```

Qy	222	GGCCATCATCTTTCACAGGGGTCCCTGTACATATGCCACAGAGGATTTAAAGGAGGCCATGGG	281
Db	428	GGCCATTTCTTTCAGGAGGCTCTGTATCAATTCAGGAGGGCCACATTCAAACATGGCCACGGG	487
Qy	282	AGTCTTTTGCAGGGTGCCTGGGAATTACTACTCCAGCTTTTCATGTTGAGCTGCATCA	341
Db	488	AGAGTTCACTGTCTCCCTCGTGTGATCAATTTTGGCTTTTGACATTCGACTGTTTCA	547
Qy	342	TTGCAAGGTGAATATTTGGCTTAATGAGGAACAAATTTTGCTTAATAGGAAGAA	396
Db	548	GAGTTCTGTGAAGATCAGGCTCATGAGGATGGTATCCAGGTCAGAGAGAAGAA	602
RESULT 5			
AB067813S3			
LOCUS			
DEFINITION Callosciurus caniceps HP-25 pseudogene, exon 3, complete cds.			
ACCESSION AB069679			
VERSION AB069679.1 GI:18149913			
KEYWORDS			
SEGMENT			
SOURCE			
ORGANISM			
Callosciurus caniceps DNA, clone:lambda TS25G1.			
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Scuridae; Sciurinae; Callosciurus.			
REFERENCE			
1			
AUTHORS			
Kojima, M., Shiba, T., Kondo, N. and Takamatsu, N.			
TITLE			
The tree squirrel HP-25 gene is a pseudogene			
JOURNAL			
Eur. J. Biochem. 268 (22), 5997-6002 (2001)			
MEDLINE			
21579740			
REFERENCE			
2 (bases 1 to 559)			
AUTHORS			
Takamatsu, N.			
TITLE			
Direct Submission			
JOURNAL			
Submitted (06-AUG-2001) Nobuhiko Takamatsu, Kitasato University, Department of Biosciences, School of Science; 1-15-1 Kitasato, Sagamihara, Kanagawa 228-8555, Japan			
(E-mail: takamatu@jet.sci.kitasato-u.ac.jp, Tel:81427789408, Fax:81427789408)			
FEATURES			
source			
1..559			
/organism="Callosciurus caniceps"			
/db_xref="taxon:64664"			
/clone="lambda TS25G1"			
join(AB067813,1:1503).1658,AB069678,1:26..185,39..559)			
/gene="HP-25"			
join(AB067813,1:1503..1658,AB069678,1:26..185,39..496)			
/gene="HP-25"			
/pseudo			
/codon_start=1			
/product="HP-25"			
39..559			
/gene="HP-25"			
/number=3			
/pseudo			
BASE COUNT			
150 a 128 c 141 g 140 t			
ORIGIN			
Query Match			
15.3%; Score 68.6; DB 10; Length 559;			
Best Local Similarity 55.5%; Pred. No. 2.6e-09;			
Matches 156; Conservative 0; Mismatches 119; Indels 6; Gaps 1;			
Qy	91	GGGCCTCTGGTCAGCCAGGTTTACCACATATACAGGAGNAATAGTGAATACACAAA	150
Db	38	GGACATCCAGGTATACCAGGAAGCCAGGATGACGAGGCCCTCTCGGAGACATGGAGCG	97
Qy	151	TGCCCTCTGCTGATAGAAAGGTTCAGCCTTTACTGTGAAGCTCAGTGGAAAACTTCCT	210
Db	98	TGCCCATCACC-----ACCAAAATCTGCCTTTGCCGTGAAGCTGAGTGAGCGTCCCCCA	151
Qy	211	CTTCCTTTCAAGGCCCATCATCTCTCACAGGGGTCTCTACATGCCACCGGATTTAAAG	270

Db	152	GGGTCTTCCAGCCCATCATATTCAGGAGTCTCTGTGTAACAACCGGAGGACCACCTATAAC	211
Qy	271	GAGGCCATGGGAGTCTTTTGGCTTGCAGGCTGCCTGGGAATTTACTACTCCAGCTTTGATGTT	330
Db	212	GTGACCTAGTGAGAGTTCAGCTATACCAACCCCGTGTGTACAAATTTTGGCTTTGACATT	271
Qy	331	GAGCTGCATCATTCGACGGTGAATATTTGGCTATATGAGAA	371
Db	272	GGACTGTTTCAGAGTCTCTGTGAAGATAAGTCTCATGAGAA	312
RESULT 6			
HS302D9/c			
LOCUS			
DEFINITION			
Human DNA sequence from clone RP1-302D9 on chromosome 22 Contains			
GSSS, complete sequence.			
ACCESSION			
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
Homo sapiens.			
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
1 (bases 1 to 145880)			
Direct Submission			
Submitted (08-DEC-1999)			
CB10 ISA, UK E-mail enquiries: humquery@sanger.ac.uk			
requests: clonerequest@sanger.ac.uk			
On Dec 13, 1999 this sequence version replaced gi:3164067.			
During sequence assembly data is compared from overlapping clones.			
Where differences are found these are annotated as variations			
together with a note of the overlapping clone name. Note that the			
variation annotation may not be found in the sequence submission			
corresponding to the overlapping clone, as we submit sequences with			
only a small overlap as described above.			
This sequence has been finished according to sequence map criteria			
as follows. An attempt is made to resolve all sequencing problems,			
such as compressions and repeats, but not necessarily within known			
annotated human repeat sequence elements (e.g. Alu). Where the			
sequence is ambiguous, there is an annotation using the 'unsure'			
feature key.			
The following abbreviations are used to associate primary accession			
numbers given in the feature table with their source databases:			
En.; EMBL; Sw.; SWISSPROT; Tr.; TREMBL; Wp.; WORMPEP; Information			
on the WORMPEP database can be found at			
http://www.sanger.ac.uk/projects/C-elegans/wormpep/			
This sequence was generated from part of bacterial clone contigs of human			
chromosome 22, constructed by the Sanger Centre Chromosome 22			
Mapping Group. Further information can be found at			
http://www.sanger.ac.uk/HGP/Chr22			
RP1-302D9 is from the library RPCI-1 constructed at the Roswell			
Park Cancer Institute by the group of Pieter de Jong. For further			
details see http://bacpac.med.buffalo.edu/			
VCTOR: pCYPAC2			
This sequence is the entire insert of clone RP1-302D9 The true left			
end of clone CTA-282F2 is at 69682 in this sequence. The true right			
end of clone CTA-415G2 is at 53167 in this sequence.			
FEATURES			
source			
1. 145880			
/organism="Homo sapiens"			
/db_xref="taxon:9606"			
/chromosome="22"			
/clone="RP1-302D9"			
/clone_lib="RPCI-1"			
188..245			
/note="MER3 repeat: matches 144..209 of consensus"			
246..571			
/note="AluX repeat: matches 1..312 of consensus"			
572..759			
/note="MER3 repeat: matches 1..144 of consensus"			
783..933			
/note="MER5A repeat: matches 26..187 of consensus"			
repeat_region			
repeat_region			
repeat_region			
repeat_region			

repeat_region	1033. .1336	/note="AluSp repeat: matches 1. .299 of consensus"	repeat_region	/note="MIR repeat: matches 76. .174 of consensus"
repeat_region	1450. .1583	/note="MIR repeat: matches 24. .160 of consensus"	repeat_region	/note="MIR repeat: matches 77. .189 of consensus"
repeat_region	1687. .1752	/note="L2 repeat: matches 2593. .2661 of consensus"	repeat_region	/note="MIR repeat: matches 24. .142 of consensus"
repeat_region	2350. .2660	/note="AluSc repeat: matches 3. .309 of consensus"	repeat_region	/note="AluY repeat: matches 1. .301 of consensus"
repeat_region	2684. .2981	/note="AluSq repeat: matches 2. .300 of consensus"	repeat_region	14368. .14452
repeat_region	3323. .3343	/note="MLTIE repeat: matches 116. .136 of consensus"	repeat_region	14589. .14679
repeat_region	3344. .3652	/note="AluY repeat: matches 1. .309 of consensus"	repeat_region	14597. .15201
repeat_region	3653. .3928	/note="MLTIE repeat: matches 136. .359 of consensus"	repeat_region	14868. .15040
repeat_region	3929. .4278	/note="THE1B repeat: matches 3. .364 of consensus"	repeat_region	15071. .15188
repeat_region	4279. .4485	/note="MLTIE repeat: matches 359. .568 of consensus"	repeat_region	15304. .15399
repeat_region	5073. .5176	/note="52 copies 2 mer ct 78 conserved"	repeat_region	15490. .15662
repeat_region	5181. .5491	/note="AluJb repeat: matches 1. .311 of consensus"	repeat_region	15669. .15727
repeat_region	6369. .6485	/note="WADE1 repeat: matches 1. .23 of consensus"	repeat_region	15728. .16027
repeat_region	6647. .6685	/note="AluX repeat: matches 1. .302 of consensus"	repeat_region	16028. .16245
repeat_region	6686. .6987	/note="WADE1 repeat: matches 23. .77 of consensus"	repeat_region	16546. .16854
repeat_region	6988. .7036	/note="AluJb repeat: matches 1. .295 of consensus"	repeat_region	18296. .18323
repeat_region	7482. .7754	/note="AluJb repeat: matches 1. .295 of consensus"	repeat_region	18324. .18392
repeat_region	7775. .8060	/note="AluJb repeat: matches 1. .295 of consensus"	repeat_region	18393. .18712
repeat_region	8414. .8551	/note="MIR repeat: matches 147. .262 of consensus"	repeat_region	18713. .19133
repeat_region	8914. .9030	/note="MIR repeat: matches 91. .262 of consensus"	repeat_region	19251. .19719
repeat_region	9110. .9280	/note="MIR repeat: matches 15. .144 of consensus"	repeat_region	19537. .20290
repeat_region	9283. .9412	/note="FAM repeat: matches 3. .161 of consensus"	repeat_region	20317. .20382
repeat_region	9521. .9679	/note="MIR repeat: matches 2. .425 of consensus"	repeat_region	20513. .20666
repeat_region	9820. .10225	/note="MIR repeat: matches 79. .150 of consensus"	repeat_region	20682. .21008
misc_feature	complement(10179. .10678)	/note="match: GSS: Em:B56592"	repeat_region	21239. .21553
misc_feature	complement(10204. .10728)	/note="match: GSS: Em:A0701486"	repeat_region	21882. .22254
misc_feature	complement(10249. .10706)	/note="match: GSS: Em:AQ225495"	repeat_region	22302. .22537
repeat_region	10312. .10383	/note="MIR repeat: matches 147. .262 of consensus"	repeat_region	22538. .22850
misc_feature	10718. .11310	/note="match: GSS: Em:B14024"	repeat_region	22851. .23801
misc_feature	10784. .11201	/note="match: GSS: Em:B43656"	repeat_region	23905. .23989
repeat_region	11838. .11946	/note="MIR repeat: matches 20. .137 of consensus"	repeat_region	23905. .23989
repeat_region	12174. .12445	/note="L2 repeat: matches 1988. .2275 of consensus"	repeat_region	23905. .23989
repeat_region	12444. .12642	/note="MIR repeat: matches 63. .241 of consensus"	repeat_region	23905. .23989
misc_feature	13017. .13369	/note="match: STS: Em:G49301"	repeat_region	23905. .23989
repeat_region	13331. .13397	/note="MIR repeat: matches 174. .244 of consensus"	repeat_region	23905. .23989
repeat_region	13398. .13698	/note="AluSp repeat: matches 1. .302 of consensus"	repeat_region	23905. .23989
repeat_region	13699. .13810		repeat_region	23905. .23989

Query Match 15.3%; Score 68.2; DB 9; Length 145880;
Best Local Similarity 58.7%; Pred.No. 5.6e-09;
Matches 118; Conservative 0; Mismatches 83; Indels 0; Gaps 0;

QY 171 AAGGTCAGCCTTTACTGTGAAGCTCAGTGAAGAACTTCTCTTCTTCAAGCCCATCAT 230

```

Db 90879 AAGATCTGCAATTTGGTGAACACTGATCGGACCCCTCCAGTGCCCTTCTAGCCCAATGTT 90820
Qy 231 CTTACAGGGTCTCTGTAACATGCCACAGGAGGATTAAGGAGGCCATGGGAGTCTTTGC 290
Db 90819 CTTCAAGGAAGCCCTTTATATACTACTCAATTCATTCCTGTTGTTGAGGATTCAC 90760
Qy 291 TTGCAGGTGCTGCTGGAATTAATCTACTCCAGCTTTGATGTTGAGCTGCATCATTCGAAGGT 350
Db 90759 CTGTACCACTCCCTAGTGTGTAACCACTTTGGTTTGAATTTGAGCTGTTTCAGAGCTGTGC 90700
Qy 351 GAATATTTGGCTAATGAGAA 371
Db 90699 AAATGTGGGTATTATGAGAA 90679

RESULT 7
HS732E4
LOCUS
DEFINITION
Human DNA sequence from clone CTA-732E4 on chromosome 22q12.1
ACCESSION
AL008722
VERSION
AL008722.16 GI:5419644
KEYWORDS
HTG.
SOURCE
Homo sapiens.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 90497)
Lloyd.D.
Direct Submission
Submitted (08-DEC-1999) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerequest@sanger.ac.uk
On Jul 8, 1999 this sequence version replaced gi:5304865.
This sequence has been finished according to sequence map criteria
as follows. An attempt is made to resolve all sequencing problems,
such as compressions and repeats, but not necessarily within known
annotated human repeat sequence elements (e.g. Alu). Where the
sequence is ambiguous, there is an annotation using the 'unsure'
feature key.
This sequence was generated from part of bacterial clone contigs of
human chromosome 22, constructed by the Sanger Centre Chromosome 22
Mapping Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr22
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
The following abbreviations are used to associate primary accession
numbers given in the feature table with their source databases:
Em., EMBL; Sw., SWISSPROT; Tr., TrEMBL; Wp., WormPEP; Information
on the WormPEP database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep
from the human BAC library described in U-J. Kim et al. (1996)
Genomics 34, 213-218.
VECTOR: pBelBAC11
This sequence is the entire insert of clone CTA-732E4 The true left
end of clone RP11-541J16 is at 9510 in this sequence. The true
right end of clone CTA-544A11 is at 41939 in this sequence.
FEATURES
source
1..90497
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="22"
/map="q12.1"
/clone="CTA-732E4"
/clone_lib="CIT978SK-A2"
repeat_region
2..631
/note="L1M1 repeat: matches 994..1640 of consensus"
repeat_region
632..757

```

```

/note="AluJb repeat: matches 1..126 of consensus"
758..1385
/note="L1P repeat: matches 3548..4173 of consensus"
1374..1869
/note="L1P repeat: matches 3221..3716 of consensus"
1952..2076
/note="L1MD2 repeat: matches 5142..5259 of consensus"
2760..3005
/note="L1MD2 repeat: matches 5259..6173 of consensus"
3418..3477
/note="30 copies 2 mer ac 90 conserved"
3423..3478
/note="14 copies 4 mer caca 94 conserved"
3670..3742
/note="MER82 repeat: matches 580..653 of consensus"
3763..4128
/note="MER82 repeat: matches 1..386 of consensus"
3879..4353
/note="match: GSS: Em:AQ559531"
4364..4569
/note="L2 repeat: matches 2535..2750 of consensus"
4612..4930
/note="AluJo repeat: matches 1..307 of consensus"
4934..5220
/note="AluSx repeat: matches 1..287 of consensus"
5221..5268
/note="16 copies 3 mer taa 97 conserved"
complement(5266..5727)
/note="match: GSS: Em:AQ028870"
complement(5282..5740)
/note="match: GSS: Em:AQ557814"
5417..5569
/note="MIR repeat: matches 2..167 of consensus"
5645..5708
/note="32 copies 2 mer at 71 conserved"
6469..6726
/note="match: STS: Em:G03694; match: STS: Em:G03694"
7573..7749
/note="MIR repeat: matches 22..207 of consensus"
10322..16495
/note="L1PA2 repeat: matches 7..6146 of consensus"
16285..16787
/note="match: GSS: Em:AQ186300"
16926..16967
/note="21 copies 2 mer aa 76 conserved"
17850..17971
/note="AluJo repeat: matches 19..147 of consensus"
17972..18285
/note="AluSg1 repeat: matches 1..309 of consensus"
18286..18472
/note="AluJo repeat: matches 147..309 of consensus"
19710..20005
/note="AluSx repeat: matches 1..296 of consensus"
20789..21262
/note="match: GSS: Em:B14443"
20811..21058
/note="match: STS: Em:L04555"
21369..21551
/note="MIR repeat: matches 2..187 of consensus"
22494..22697
/note="MIR repeat: matches 35..236 of consensus"
23443..23755
/note="AluSx repeat: matches 1..312 of consensus"
24052..24348
/note="AluSx repeat: matches 1..295 of consensus"
25623..25670
/note="24 copies 2 mer ac 77 conserved"
26003..26192
/note="MIR repeat: matches 20..216 of consensus"
27067..27138
/note="Charlie4 repeat: matches 1902..1961 of consensus"
27363..27661
/note="AluYb8 repeat: matches 1..306 of consensus"

```



```

/misc_feature  /note="assembly_fragment"
48699. .52567
/note="assembly_fragment"
52668. .55272
/note="assembly_fragment"
clone_end:SP6
vector_side:left"
55373. .59485
/misc_feature  /note="assembly_fragment"
59586. .63994
/misc_feature  /note="assembly_fragment"

```

Query Match 14.1%; Score 63.2; DB 2; Length 144514;
 Best Local Similarity 74.1%; Pred. NO. 1.7e-07;
 Matches 80; Conservative 0; Mismatches 28; Indels 0; Gaps 0;

```

QY 340 CATTCAAGTCAATATTGGCTAATGAGCAAGCAAAATTTGGCTAATAAGCAAGAAATTT 399
|||||
Db 133169 CCTAGAGATTTGTGACCTTTGAACCTGGAAGAGATGATTTAGAGTATCAGGGAAGAAATTT 133228
|||||
QY 400 TCTAAGCAGCAAGCAATTCAGAGGCTGACTTGGGTGCTGCTTAAAGGCA 447
|||||
Db 133229 TCTAAGCAGCAAGCAATTCAGAGGCTGACTTGGGTGCTGCTTAAAGCA 133276
|||||

```

```

RESULT 11
AC025112      149038 bp  DNA  linear  HTG 08-APR-2000
LOCUS
DEFINITION
Homosapiens chromosome 6 clone RP11-608N7 map 6, WORKING DRAFT
SEQUENCE, 27 unordered pieces.
AC025112
AC025112.2 GI:7528131
VERSION
KEYWORDS
HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE
Homosapiens.

```

```

ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Birken,B., Linton,L., Nussbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
Boguslavskiy,L., Boukhgalter,B., Brown,A., Burkett,G.,
Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
Collimore,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S.,
Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
Klein,J., LaRoque,K., Lamazares,R., Landers,T., Lehoczy,J.,
Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
Melridm,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
O'Neill,D., Olivari,T.M., Oliver,J., Peterson,K., Pierre,N.,
Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigglio,J.,
Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
Young,G., Zainoun,J., Zimmer,A. and Zody,M.
Direct Submission
Submitted (05-MAR-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Apr 8, 2000 this sequence version replaced gi:7158923.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR

```

```

TITLE
JOURNAL
COMMENT

```

```

Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L6882
Center clone name: 608_N_7
----- Summary Statistics
Sequencing vector: M13; M7815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 132985 bases at least Q40
Consensus quality: 140748 bases at least Q30
Consensus quality: 143894 bases at least Q20
Insert size: 138000; agarose-fp
Insert size: 146438; sum-of-contigs
Quality coverage: 3.6 in Q20 bases; agarose-fp
Quality coverage: 3.4 in Q20 bases; sum-of-contigs
-----

```

* NOTE: This is a 'working draft' sequence. It currently consists of 27 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence, as soon as it is available and the accession number will be preserved.

```

* 1 165: contig of 165 bp in length
* 166 265: gap of 100 bp
* 266 1355: contig of 1050 bp in length
* 1356 1455: gap of 100 bp
* 1456 1752: contig of 297 bp in length
* 1753 1852: gap of 100 bp
* 1853 3336: contig of 1484 bp in length
* 3337 3436: gap of 100 bp
* 3437 4814: contig of 1378 bp in length
* 4815 4914: gap of 100 bp
* 4915 6853: contig of 1939 bp in length
* 6854 6953: gap of 100 bp
* 6954 9114: contig of 2161 bp in length
* 9115 9214: gap of 100 bp
* 9215 12071: contig of 2857 bp in length
* 12072 12171: gap of 100 bp
* 12172 14633: contig of 2462 bp in length
* 14634 14733: gap of 100 bp
* 14734 17354: contig of 2621 bp in length
* 17355 17454: gap of 100 bp
* 17455 21910: contig of 4456 bp in length
* 21911 22010: gap of 100 bp
* 22011 26287: contig of 4277 bp in length
* 26288 26387: gap of 100 bp
* 26388 31094: contig of 4707 bp in length
* 31095 31194: gap of 100 bp
* 31195 35408: contig of 4214 bp in length
* 35409 35508: gap of 100 bp
* 35509 40764: contig of 5256 bp in length
* 40765 40864: gap of 100 bp
* 40865 45639: contig of 4775 bp in length
* 45640 45739: gap of 100 bp
* 45740 50826: contig of 5087 bp in length
* 50827 50926: gap of 100 bp
* 50927 56617: contig of 5691 bp in length
* 56618 56717: gap of 100 bp
* 56718 63406: contig of 6689 bp in length
* 63407 63506: gap of 100 bp
* 63507 71321: contig of 7815 bp in length
* 71322 71421: gap of 100 bp
* 71422 81641: contig of 10220 bp in length
* 81642 81741: gap of 100 bp
* 81742 91982: contig of 10241 bp in length
* 91983 92082: gap of 100 bp
* 92083 102357: contig of 10275 bp in length
* 102358 102457: gap of 100 bp
* 102458 112952: contig of 10495 bp in length
* 112953 113052: gap of 100 bp

```

* 113053 124237: contig of 11185 bp in length
 * 124238 124337: gap of 100 bp
 * 124338 135694: contig of 11357 bp in length
 * 135695 135794: gap of 100 bp
 * 135795 149038: contig of 13244 bp in length.

FEATURES

```

source
1. .149038
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="6"
/map="6"
/clone="RP11-608N7"
/clone_lib="RPCI-11 Human Male BAC"
1. .165

misc_feature
1. .165
/note="assembly_fragment"
clone_end:T7
vector_side:right"

misc_feature
266. .1355
/note="assembly_fragment"

misc_feature
1456. .1752
/note="assembly_fragment"
clone_end:SP6
vector_side:left"

misc_feature
1853. .3336
/note="assembly_fragment"

misc_feature
3437. .4814
/note="assembly_fragment"

misc_feature
4915. .6853
/note="assembly_fragment"

misc_feature
6954. .9114
/note="assembly_fragment"

misc_feature
9215. .12071
/note="assembly_fragment"

misc_feature
12172. .14633
/note="assembly_fragment"

misc_feature
14734. .17354
/note="assembly_fragment"

misc_feature
17455. .21910
/note="assembly_fragment"

misc_feature
22011. .26287
/note="assembly_fragment"

misc_feature
26388. .31094
/note="assembly_fragment"

misc_feature
31195. .35408
/note="assembly_fragment"

misc_feature
35509. .40764
/note="assembly_fragment"

misc_feature
40865. .45639
/note="assembly_fragment"

misc_feature
45740. .50826
/note="assembly_fragment"

misc_feature
50927. .56617
/note="assembly_fragment"

misc_feature
56718. .63406
/note="assembly_fragment"

misc_feature
63507. .71321
/note="assembly_fragment"

misc_feature
71422. .81641
/note="assembly_fragment"

misc_feature
81742. .91982
/note="assembly_fragment"

misc_feature
92083. .102357
/note="assembly_fragment"

misc_feature
102458. .112952
/note="assembly_fragment"

misc_feature
113053. .124237
/note="assembly_fragment"

misc_feature
124338. .135694
/note="assembly_fragment"

misc_feature
135795. .149038
/note="assembly_fragment"

BASE COUNT 44765 a 26643 c 27220 g 47805 t 2605 others
ORIGIN

```

```

Query Match 14.1%; Score 63.2; DB 2; Length 149038;
Best Local Similarity 74.1%; Pred.No. 1.7e-07;
Matches 80; Conservative 0; Mismatches 28; Indels 0; Gaps 0;

QY 340 CATTGCAAGTGAATATTGGCTAATGAGGACAAATTTGGCTAATAAGGAAGAATT 399
Db 90741 CCTAGAGATTGTGAACCTTTGAACCTGAAAGAGATGATTAGAGTATCAGGGAAGAATT 90800
|||||
QY 400 TCTAAGCAGCAAAAGCATTCAAGAGGTGACTTGGTCTGTTAAAGGCA 447
|||||
Db 90801 TCTAAGCAGCAAAAGCATTCAAGAGGTGACTTGGTCTGTTAAAGGCA 90848
|||||

```

```

RESULT 12
AC015553/c
LOCUS
DEFINITION
Homo sapiens 9 BAC RP11-100N10 (Roswell Park Cancer Institute Human
BAC Library) complete sequence.
AC015553
AC015553.21 GI:13489133
HTG.
SOURCE
Homo sapiens.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 162063)
Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Banks,T., Barbara,J.,
Benton,J., Bimaga,K., Blankenburg,K., Bonnin,D., Bouck,J.,
Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,
Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carlton,T.F.,
Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,
Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,
Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,
Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,
Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,
Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,
Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A.,
Hernandez,J., Hernandez,O., Hodgson,A., Hognes,M., Holloway,C.,
Hollins,B., Honsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J.,
Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,
Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,
Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,
Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
Louiseged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,
Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P., Meador,M.,
Mei,G., Metzker,M., Miner,G., Moser,M., Neal,D., Newton,J., Newton,N.,
Morgan,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S.,
Oguh,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B.,
Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.,
Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojubokan,I., Rolfe,M.,
Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shooshtari,N.,
Sisson,I., Sodergren,E., Sonaike,T., Sparks,A., Stanley,H.,
Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K.,
Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N.,
Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R.,
Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C.,
Watlington,S., Williams,G., Williamson,A., Wleczek,R., Wooden,S.,
Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
and Gibbs,R.
Direct Submission
Unpublished
2 (bases 1 to 162063)
Worley,K.C.
Direct Submission
Submitted (17-NOV-1999) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One

```

```

TITLE
JOURNAL
REFERENCE
2 (bases 1 to 162063)
Worley,K.C.
Direct Submission
TITLE
JOURNAL

```

REFERENCE

AUTHORS Baylor Plaza, Houston, TX 77030, USA

TITLE 3 (bases 1 to 162063)

JOURNAL Direct Submission

Submitted (30-MAR-2001) Human Genome Sequencing Center, Department

of Molecular and Human Genetics, Baylor College of Medicine, One

Baylor Plaza, Houston, TX 77030, USA

4 (bases 1 to 162063)

AUTHORS Worley, K.C.

TITLE Direct Submission

Submitted (12-JUL-2001) Human Genome Sequencing Center, Department

of Molecular and Human Genetics, Baylor College of Medicine, One

Baylor Plaza, Houston, TX 77030, USA

5 (bases 1 to 162063)

AUTHORS Worley, K.C.

TITLE Direct Submission

Submitted (13-JUL-2001) Human Genome Sequencing Center, Department

of Molecular and Human Genetics, Baylor College of Medicine, One

Baylor Plaza, Houston, TX 77030, USA

On Mar 30, 2001 this sequence version replaced gi:13487890.

INFORMATION: <http://www.hgsc.bcm.tmc.edu/> or email

gc-help@bcm.tmc.edu

COMMENT

CLONE LENGTH: This sequence does not necessarily represent the entire insert of this clone. Overlapping regions of clones are only sequenced and submitted once, so the sequence for the remainder of the insert may be found in the record for the adjacent clones. Overlapping clones are noted at the beginning and end of the Features listing.

ANNOTATION OF FEATURES:

STSs are identified using ePCR (Genome Res. 7:541-550) searches of a local database that includes entries from dbSTS, GDB, and local mapping efforts.

Repeats are identified using RepeatMasker (A. Smit and P. Green, unpublished.) for Human and Mouse sequences.

Genes and Region of sequence similarity are identified by BLAST (Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the EST and cDNA sequences. Genes demonstrate at least two exons flanked by consensus splice sites that maintained sequence continuity across the splice junctions. Sequences that are not identical matches are annotated as similar.

SEQUENCING READ COVERAGE: Sequencing is completed to a minimum standard of double strand coverage with a minimum of 2 clones and 2 reads with no ambiguities or 2 chemistries with a minimum of 2 clones and 3 reads with no ambiguities. If the sequence quality for a region does not meet this standard, it will be indicated in the annotation as Low Coverage.

QUALITY OF INDIVIDUAL BASES: This sequence meets stringent quality standards - estimated error rate less than 1 per 10,000 bases. Reports of lowest quality individual bases and measures of base quality are listed below. Description of the metrics can be found at URL: <http://gc.bcm.tmc.edu:8088/quality.info/genbank.annotation.html>.

QUALSTAT-REPORT-----

```

----- Summary Statistics -----
Contig length: 162063
Phrap values in estimate: 161326
Average error rate (BCM-Phrap estimate): 9.86316e-06
Fraction of Phrap values less than 40 : 0.00234308
Number of consensus changing edits: 18
Number of N's in consensus : 0

```

```

----- Consensus changing edits -----
Position Original+Context Edited+Context
10095 caccacatt(t)acctccaaa
10096 acccaccitt(a)ccctccaaa

```

```

37628 aggaaggaag(n)aaggaaggaag
41530 ccagcaataa(n)caagtgaat
68252 gctggaattg(n)tccttattt
68273 catctagtg(n)nggtccatg
68274 atctagtg(n)ggttccatg
68288 tccatttg(n)atacatagat
87792 tggtagaag(n)caacagaaga
98412 acgtactta(n)ncagggtccc
113818 cgttaacttan(n)caaggtccc
113821 tgatactgaa(n)ccnctaaaaa
113849 tactgaacc(n)ctaataaaaa
122621 aaaaaaaa(n)gactaaaaata
133372 atggaccatg(n)tgctttgtga
134847 gttataaaa(n)caataaacta
160961 tctttactg(n)ggcctatta
          aaaaaaaat(c)caaaaacttg

```

----- Distribution of Quality < 40 Bases -----

```

500|
450|
400|
350|
300|
250|
200|
150|
100|
 50|
  0|
      * * * * *
      5 10 15 20 25 30 35 40
      Phrap Value Range

```

Version: 1.01 qxfo.

FEATURES Location/Qualifiers

1..162063

/organism="Homo sapiens"

/db_xref="taxon:9606"

/chromosome="9"

/clone="RP11-100N10"

706..734

/rpt_family="AT-rich"

complement(738..930)

/rpt_family="L1M4"

complement(943..1086)

/rpt_family="L1ME3A"

1979..2255

/rpt_family="L1M8"

2560..2722

/rpt_family="AluJb"

2723..2831

/rpt_family="GA-rich"

2832..2859

/rpt_family="(GAAA)n"

2871..3150

/rpt_family="AluY"

3151..3335

/rpt_family="(GAAA)n"

3336..3346

/rpt_family="AluY"

3347..3481

/rpt_family="L1M4"

3482..4108

/rpt_family="L1M4"

4109..4207

/rpt_family="L1M4"

Query Match

Best Local Similarity

14.0%; Score 62.6; DB 9; Length 162063;

79.6%; Pred. No. 2.6e-07;


```
mat_peptide 155..709
/product="hp-27"
polyA_signal 1369..1374
polyA_site 1385
BASE COUNT 439 a 285 c 311 g 350 t
ORIGIN
Query Match 14.0%; Score 62.4; DB 10; Length 1385;
Best Local Similarity 55.6%; Pred. No. 2e-07;
Matches 120; Conservative 0; Mismatches 96; Indels 0; Gaps 0;
QY 156 CTGTCCTGATATAGAAAGGTCAGCCCTTTACTGTGAAGCTCAGTGGGAAAACCTCCTCTTCC 215
Db 313 CTGCCACAGCAAGGAACATCGGCCCTTTGCAGTGAAGGCAAAATGAGCTGCCCCCACTCC 372
QY 216 TTTCAAGCCCATCATCTTCACAGGGGTCTGTACAATGCCAGAGGGATTAAAGGAGGC 275
Db 373 CTCCCAGCCCGTGATCTTCAAGGAAGCCCTGATGACGCTCAGGACACTTTGATCTGGC 432
QY 276 CATGGGAGTCTTTGCTGTCAGGGTGCCTGGGAATTACTACTCAGCTTTGATGTTGAGCT 335
Db 433 CACTGTGTGTTTACCTGCCAGTCCCAAGGACTCTTACCAGTTTGGATTTCACATTGAAGC 492
QY 336 GCATCATTTGCAAGGTGAATATTGGCTAATGAGGAA 371
Db 493 TGTCCAGAGGGGTGTGAAGGTGAGCCTCATGAGAA 528
```

Search completed: February 20, 2003, 05:05:34
Job time : 2438.16 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 19, 2003, 23:48:42 ; Search time 391.859 seconds
(without alignments)
6406.126 Million cell updates/sec

Title: us-09-997-610-1-copy_2_156

Perfect score: 155

Sequence: 1 atagtggtacatacctgtctt.....agtgaatgacaaaatgcc 155

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*

1: em_estba:*

2: em_esthum:*

3: em_estin:*

4: em_estmu:*

5: em_estov:*

6: em_estpl:*

7: em_estro:*

8: em_htc:*

9: gb_estl:*

10: gb_est2:*

11: gb_htc:*

12: gb_est3:*

13: gb_est4:*

14: gb_est5:*

15: em_estfun:*

16: em_estom:*

17: gb_gss:*

18: em_gss_hum:*

19: em_gss_inv:*

20: em_gss_pln:*

21: em_gss_vrt:*

22: em_gss_fun:*

23: em_gss_mam:*

24: em_gss_mus:*

25: em_gss_other:*

26: em_gss_pro:*

27: em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	39.6	25.5	342	14	244751 HSC28C121 n
2	39.4	25.4	638	9	AU205784 AU205784
3	38.6	24.9	483	13	BI322379 kx20d10.y
c	38.6	24.9	531	10	BE121287 UI-R-CAO-
5	38.6	24.9	552	13	BI863711 kx47e02.y
6	37.8	24.4	352	9	AI617644 zehn1908.

7	37.8	24.4	770	13	BJ121588
8	37.4	24.1	468	10	BE580165
9	37.4	24.1	889	14	BQ939199
c	37.2	24.0	651	17	CNS05LPT
11	37.2	24.0	1101	17	CNS04XPG
12	37	23.9	454	10	BE581564
13	36.8	23.7	525	10	AM675997
14	36.8	23.7	575	10	AM179710
15	36.8	23.7	585	10	AM179840
16	36.8	23.7	605	10	AM347957
17	36.8	23.7	624	12	BE758489
18	36.8	23.7	640	10	AM179745
19	36.8	23.7	647	10	AM409404
20	36.8	23.7	647	10	AM409406
21	36.8	23.7	648	10	AM874708
22	36.8	23.7	654	10	AM348036
23	36.8	23.7	695	10	BE285930
24	36.8	23.7	701	10	AM179849
25	36.8	23.7	776	13	BI554667
26	36.8	23.7	801	9	AA585675
27	36.2	23.4	360	14	D69730
28	36.2	23.4	407	10	BE581103
29	36.2	23.4	420	10	BE579707
30	36.2	23.4	441	10	BE581944
31	36.2	23.4	478	10	BE579460
32	36.2	23.4	491	10	BE579541
33	36.2	23.4	492	10	BE579271
34	36.2	23.4	496	10	BE579572
35	36.2	23.4	632	13	BJ136441
c	36.2	23.4	726	13	BJ108893
37	36	23.2	450	13	BJ002073
38	36	23.2	461	13	BI323061
39	36	23.2	865	14	BQ721294
40	35.8	23.1	407	13	BI742359
41	35.8	23.1	458	14	BQ35482
42	35.8	23.1	467	13	BI323796
43	35.8	23.1	483	10	BE579244
44	35.8	23.1	675	12	BG310518
45	35.8	23.1	786	10	BE383449

ALIGNMENTS

RESULT 1
244751
LOCUS 342 bp mRNA linear EST 14-NOV-1994
DEFINITION HSC28C121 normalized infant brain cDNA Homo sapiens cdna clone
C-28C12, mRNA sequence.
ACCESSION 244751
VERSION 244751.1 GI:573911
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 342)
AUTHORS Auffray,C., Behar,G., Bois,F., Bouchier,C., da Silva,C., Devignes,M.D., Duprat,S., Houlgatte,R., Jumeau,M.N., Lamy,B., Lorenzo,F., Mitchell,H., Mariage-Samson,R., Pietu,G., Pouliot,Y., Sebastiani-Kabaktchis,C. and Tessier,A.
TITLE IMAGE: molecular integration of the analysis of the human genome and its expression
JOURNAL C.R.Acad. Sci. III, Sci. Vie 318 (2), 263-272 (1995)
MEDLINE 93277534
COMMENT Contact: Genethon
Genexpress-Genethon
Genethon Centre de recherche sur le Genome Humain
1, rue de l'Internationale, BP60 91002 EVRY Cedex, FRANCE
Tel: 33169472800
Fax: 33160778698
Email: genexpress@genethon.fr
Single read.

```
Genexpress_library_id: C; Genexpress_sequence_id: ylc-28cl2
Seq primer: (-21)JM3_universal.
Location/Qualifiers
  source
    1..342
      /organism="Homo sapiens"
      /db_xref="taxon:9606"
      /clone_lib="c-28cl2"
      /clone_lib="normalized infant brain cDNA"
      /sex="female"
      /tissue_type="total brain"
      /dev_stage="3 months old"
      /note="Organ: brain; Vector: lafmid BA; Site_1: HindIII;
      Site_2: NotI; sex=female; dev_stage=3 months old;
      isolate=muscular atrophy patient; tissue_type=total brain
      ; total mRNA was oligo-(dT) primed and directionally
      cloned 5' -> 3' into the HindIII -> NotI sites of the
      lafmid BA vector. Clone library from B.Soures, Psychiatry
      Dept. Columbia University, USA. Normalization_method:
      Bento Soares, P.N.A.S in press"
BASE COUNT      78 a 79 c 89 g 88 t      8 others
ORIGIN
Query Match      25.5%; Score 39.6; DB 14; Length 342;
Best Local Similarity 55.1%; Pred. No. 1;
Matches 75; Conservative 0; Mismatches 61; Indels 0; Gaps 0;

QY 13 CCGTCTTAATAACGCGCAGTCATTGACGATGTAGAAGTTGTGGACCTCCAGCACACCC 72
Db 48 CCGTCCGACTCTCTCGCGGTGTGTCAGTGTGAATTTGTACAGCTGCATCTGCTCCC 107

QY 73 AGGCCCCCAGAGAAGTGCGGCTCCTGCTGTGCACCGAGTTTACCACAATATACAGGAGAA 132
Db 108 CGCCCCGAGAGCCGAGGAGCTACCGTGGCTCCNGANTTTCCCAAAATTCAGCAGCA 167

QY 133 ATAAAGTGAATGACAA 148
Db 168 GAAATTCAGTGACTA 183

RESULT 2
A0205784
LOCUS      A0205784      638 bp      mRNA      linear      EST 17-JUL-2001
DEFINITION      Caenorhabditis elegans cDNA library, stage L1
ACCESSION      A0205784
VERSION      A0205784.1 GI:14838413
KEYWORDS      EST.
SOURCE      Caenorhabditis elegans.
ORGANISM      Caenorhabditis elegans
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea
; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE      1 (bases 1 to 638)
AUTHORS      Kohara,Y., Shin-I,T., Thierry-Mieg,J., Thierry-Mieg,D., Suzuki,I.Y.
and Sugano,S.
TITLE      A complementary view of the C.elegans genome
JOURNAL      Unpublished (2001)
COMMENT      Contact: Yuji Kohara
Genome Biology Lab.
National Institute of Genetics
Yata 1111, Mishima, Shizuoka 411, Japan
Tel: 81-559-81-6854
Fax: 81-559-81-6855
Email: ykohara@lab.nig.ac.jp.
Location/Qualifiers
  source
    1..638
      /organism="Caenorhabditis elegans"
      /strain="N2"
      /db_xref="taxon:6239"
      /clone_lib="yk85b02"
      /clone_lib="unpublished oligo-capped cDNA library, stage
      L1"
      /sex="Hermaphrodite"
      /tissue_type="whole animal"
```

```
/dev_stage="L1"
/note="The AD-wrmcDNA library was generated with poly(A)+
RNA isolated from both hermaphrodite and male N2 worms of
all larval stages, embryos, adults and dauers and the
subsequent generation of cDNAs by poly(A) priming. The
cDNAs were cloned into pPC86"
BASE COUNT      193 a 176 c 190 g 79 t
ORIGIN
Query Match      25.4%; Score 39.4; DB 9; Length 638;
Best Local Similarity 56.6%; Pred. No. 1.2;
Matches 73; Conservative 0; Mismatches 56; Indels 0; Gaps 0;

QY 21 AATAACGGCAGTCATTAGCATGTAGAGTTGCTGCACCTCCAGCACACCCAGGCCCC 80
Db 504 ANTGAAGGAGACTCTGGCCCTCCAGGACCACAGGACTCCAGGACACCCAGGTGTCC 563

QY 81 AGAAGAAGTGGGCGCTCTGTTGGTCCAGCGTTTACACAAATATACAGGAGAAATAGTGA 140
Db 564 AGGAGACAAGGAGATTCCGAGGTGTCCAGGATTACGAGGATTCCAGGACCAAGGGAGA 623

QY 141 ANTGACAAA 149
Db 624 TGTCCGAAA 632

RESULT 3
BI322379
LOCUS      BI322379      483 bp      mRNA      linear      EST 30-JUL-2001
DEFINITION      kx20d10.y3 Parastrongyloides trichosuri FL pAMP1 vl Chiapelli
McCarter Parastrongyloides trichosuri cDNA 5' similar to TR:Q61436
Q61436 PROCOLLAGEN, TYPE IV, ALPHA 5 ; contains element PTR5
repetitive element ;, mRNA sequence.
ACCESSION      BI322379
VERSION      BI322379.1 GI:15001565
KEYWORDS      EST.
SOURCE      Parastrongyloides trichosuri.
ORGANISM      Parastrongyloides trichosuri
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
Panagrolaimoidea; Strongyloidea; Parastrongyloidea.
REFERENCE      1 (bases 1 to 483)
AUTHORS      McCarter,J., Clifton,S., Chiapelli,B., Pape,D., Martin,J., Wylie,T.
, Dante,M., Marra,M., Hillier,L., Kucaba,T., Theising,B., Bowers,Y.
, Gibbons,M., Ritter,E., Bennett,J., Franklin,C., Tsagarisvill,I., R.
, Ronko,I., Kennedy,S., Maguire,L., Beck,C., Underwood,K., Stepcoe
, M., Allen,M., Person,B., Swaller,T., Harvey,N., Schurk,R., Kohn,S.
, Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R. and
Wilson,R.
TITLE      The Washington Univ. Nematode EST Project, 1999
JOURNAL      Unpublished (1999)
COMMENT      Contact: McCarter JP
The Washington Univ. Nematode EST Project, 1999
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
The library was constructed by Brandi Chiapelli and Dr. James
McCarter (bchiapeli@watson.wustl.edu & jmccarter@watson.wustl.edu) at
Washington University, St. Louis. DNA Sequencing by: Washington
University Genome Sequencing Center St. Louis.
Seq primer: -40RP from Gibco
High quality sequence stop: 428.
Location/Qualifiers
  source
    1..483
      /organism="Parastrongyloides trichosuri"
      /db_xref="taxon:131310"
      /clone_lib="Parastrongyloides trichosuri FL pAMP1 vl
      Chiapelli McCarter"
      /dev_stage="Free Living"
      /lab_host="DH10B"
      /note="Vector: pAMP1 (Gibco); Site_1: NotI; Site_2: SalI;
      The library was constructed by Brandi Chiapelli and Dr.
```

TAG_TISSUE=corpus-striatum

BASE COUNT	124 a	134 c	128 a	145 t
124 a	134 c	128 a	145 t	

Query Match	24.9%	Score 38.6;	DB 10;	Length 531;
Best Local Similarity	59.6%	Pred. No. 1.9;		
Matches 65;	Conservative	0;	Mismatches 44;	Indels 0;
				Gaps 0;

Qy 40 CATGTAGAACTTGCTGGACCTCCAGGCACACCAGGCCCGCCAGAGAAGTGGGGCCTCCT 99
- - - - -
Dp 526 CTTTCNTGGCCTCACAGGACACTCAGGGCTCCCTGGCCCTCCCTGGACAGCAGGGGATCCT 467

[illegible]

RESULT	5
BH63711	
LOCUS	
DEFINITION	
<hr/>	
BH63711	552 bp mRNA linear EST 10-OCT-2001
MCCarter Paratrongsylodes trichosuri FL pampi Vl Chapelli	
KcCarter Paratrongsylodes trichosuri CDNA 5' similar to	
SX:CA24_ASCSU P27393 COLLAGEN ALPHA 2(IV) CHAIN PRECURSOR.	[1]
:contains element PTR5 repetitive element ; mrna sequence.	

ACCESSION BT863711

VERSION	KEYWORDS	SOURCE	ORGANISM	REFERENCE	AUTHORS
B1863/11.1	EST.	GI:16005501	Parastromyloides trichosuri.	Parastromyloides trichosuri.	McCarter, J., Clifford, S., Chiappelli, B., Pape, D., Martin, J., Wylie, T., Dante, M., Marra, M., Hillier, L., Kucaba, T., Thelsing, B., Bowers, Y., Gibbons, M., Ritter, E., Bennett, J., Franklin, C., Tsagarishevili, R., Ronko, I., Kennedy, S., Maguire, L., Beck, C., Underwood, K., Steptoe, M., Allen, M., Person, B., Swallier, T., Harvey, N., Schurk, R., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R., Waterston, R., and Wilson, R.

TITLE The Washington Univ. Nematode EST Project, 1999
JOURNAL Unpublished (1999)
COMMENT Contact: McCarter JP

Contact: McCarter JP
 The Washington Univ. Nematode EST Project, 1999
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@waton.wustl.edu
 The library was constructed by Brandi Chiapelli and Dr. James
 McCarter (bchiapelli@waton.wustl.edu & jmccarter@waton.wustl.edu) at
 Washington University, St. Louis. DNA Sequencing by: Washington
 University Genome Sequencing Center St. Louis.
 Seq primer: -40RP from Gibco
 High quality sequence spot: 395.

FEATURES
source

```

1. 552
location/qualifiers
/organism="Parastrongyloides trichosuri"
/db_xref="taxon:131310"
/clone.lib="Parastrongyloides trichosuri FL pAMPL v1
Chiapelli McCarter"
/dev_stage="Free Living"
/lab_host="Dhl08"
/notes-vector: pAMPL (Gibco); Site_1: NotI; Site_2:
The library was constructed by Brandi Chiapelli and
James McCarter at Washington University, St. Louis.
cDNA was made by using Dynabead oligo-dT priming (D
PCR based library using a modified protocol from th

```



```

ACCESSION      BE580165
VERSION        BE580165.1  GI:98311107
KEYWORDS
SOURCE
ORGANISM       Strongyloides stercoralis.
REFERENCE
AUTHORS        McCarter,J., Clifton,S., Chiapelli,B., Pape,D., Martin,J., Wyllie,T.,
               Dante,M., Marra,M., Hillier,L., Kucaba,T., Theising,B., Bowers,Y.,
               Gibbons,M., Ritter,E., Bennett,J., Franklin,C., Tsagareishvili,R.,
               Ronko,I., Kennedy,S., Maguire,L., Beck,C., Underwood,K., Steptoe
               M., Allen,M., Person,B., Swaller,T., Harvey,N., Schurk,R., Kohn,S.,
               Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R. and
               Wilson,R.
TITLE          The Washington Univ. Nematode EST Project, 1999
JOURNAL
COMMENT        Unpublished (1999)
               Contact: McCarter JP
               The Washington Univ. Nematode EST Project, 1999
               Washington University School of Medicine
               4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
               Tel: 314 286 1800
               Fax: 314 286 1810
               Email: est@watson.wustl.edu
               The library was constructed by Dr. Thomas Nutman and colleagues of
               NIAID, NIH (nutman@nih.gov). DNA Sequencing by: Washington
               University Genome Sequencing Center St. Louis.
               High quality sequence stop: 325.
               Location/Qualifiers
               1..468
               /organism="Strongyloides stercoralis"
               /strain="Rhabditiform larvae obtained from gerbills"
               /db_xref="taxon:6248"
               /clone_lib="TBN95TM-SSR"
               /lab_host="XL-1 Blue MRF" (Stratagene)
               /note="Vector: Lambda Uni-ZAP XR (Stratagene); Site_1:
               EcoRI; Site_2: XhoI; mRNA was purified from 2 x 10B3
               rhabditiform larvae which had been isolated from gerbills
               experimentally infected with larvae originally isolated
               from experimentally infected dogs. cDNA was constructed
               and, using adaptors, was cloned unidirectionally into the
               vector from the EcoRI site to the XhoI site. The library
               has an unamplified titer of 1 x 10E5 pfu/ml and an
               amplified, undiluted titer of 9 x 10E11 pfu/ml. The
               average insert size of the unamplified library is 675 bp
               (range, 100-1700)."
               157 a 92 c 134 g 85 t
               BASE COUNT 157 a 92 c 134 g 85 t
               ORIGIN
               Query Match 24.1%; Score 37.4; DB 10; Length 468;
               Best Local Similarity 57.1%; Pred. No. 4;
               Matches 68; Conservative 0; Mismatches 51; Indels 0; Gaps 0;
               Oy 24 AACGCCATGTCATGAGCATGTGAGAGTGTGGACCTCAGCACACCCAGGCCCCAGCA 83
               || || || || || || || || || || || || || || || || || || || || ||
               Db 126 AAGAGGAGAAAAGACACAAAAGAGAGACAGAGCACCAGGACAGATGGATCTCCAGG 185
               Oy 84 AGAAGTGGGGCTCTCGTGGCACCAGGTTTACCACATATACAGGAGAAATAGTGAAA 142
               || || || || || || || || || || || || || || || || || || || || ||
               Db 186 AGAACAGGACAGAGGTGAACAGGTTTACAGGAGATATTGGACCATCAGGAGAAA 244
               RESULT 9
               LOCUS      B0939199 889 bp mRNA linear EST 21-AUG-2002
               DEFINITION AGENCOURT_8922132 NCI_CGAP_CO24 Mus musculus cDNA clone
               IMAGE:6395968 5', mRNA sequence.
               ACCESSION B0939199
               VERSION B0939199
               KEYWORDS EST.
               SOURCE house mouse.
               ORGANISM Mus musculus
               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE      1 (bases 1 to 889)
AUTHORS        NIH-MGC http://mgc.nci.nih.gov/.
TITLE          National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL        Unpublished (1999)
COMMENT        Contact: Robert Strausberg, Ph.D.
               Email: cgabbs-r@mail.nih.gov
               Tissue procurement: The Cepko Laboratory
               cDNA Library Preparation: Life Technologies, Inc.
               DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
               DNA Sequencing by: Agencourt Bioscience Corporation
               Clone distribution: MGC clone distribution information can be
               found through the I.M.A.G.E. Consortium/LLNL at:
               http://image.llnl.gov
               plate: LLAM13891 row: h column: 17
               High quality sequence stop: 672.
               Location/Qualifiers
               1..889
               /organism="Mus musculus"
               /strain="FVB/N"
               /db_xref="taxon:10090"
               /clone="IMAGE:6395968"
               /clone_lib="NCI_CGAP_CO24"
               /lab_host="DH10B (T1 phage-resistant)"
               /note="organ: colon; Vector: pCMV-SPORT6; Site_1: NotI;
               Site_2: SalI; Cloned unidirectionally. Primer: Oligo dt.
               Average insert size 1.6 kb. Constructed by Life
               Technologies. Note: this is a NCI_CGAP Library."
               202 a 251 c 292 g 140 t 4 others
               BASE COUNT 202 a 251 c 292 g 140 t 4 others
               ORIGIN
               Query Match 24.1%; Score 37.4; DB 14; Length 889;
               Best Local Similarity 61.5%; Pred. No. 4.1;
               Matches 59; Conservative 0; Mismatches 37; Indels 0; Gaps 0;
               Oy 53 CTGGACCTCCACACACCCAGGCCCCAGAGAGAGTGGGGCTCTCTGTGCACACAGGTT 112
               || || || || || || || || || || || || || || || || || || || || ||
               Db 600 CAGGACAGCTCAGCGCTCCCTGGCTTCTCTGGACAGCAGGGGACACCTGGATGCCAGGTT 659
               Oy 113 TACCACATATACAGAGAAATAGTGAATGACAA 148
               || || || || || || || || || || || || || || || || || || || || ||
               Db 660 TCCCAGGTTCTANAGTGAAATGGGTGCATGGGAA 695
               RESULT 10
               LOCUS      CNS05LPT/c 651 bp DNA linear GSS 26-MAY-2000
               DEFINITION Tetraodon nigroviridis genome survey sequence T7 end of clone
               048C07 of library A from Tetraodon nigroviridis, genomic survey
               sequence.
               ACCESSION AL343082
               VERSION AL343082.1 GI:8236840
               KEYWORDS GSS: genome survey sequence.
               SOURCE Tetraodon nigroviridis.
               ORGANISM Tetraodon nigroviridis
               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
               Actinopterygii; Neopterygii; Teleostei; Euteleostei; Notoleosteoi;
               Acanthomorphi; Acanthopterygii; Percomorpha; Tetraodontiformes;
               Tetraodontidae; Tetraodon.
               1 (bases 1 to 651)
               Roest-Crolius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
               Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
               Saurin,W. and Weissenbach,J.
               Human gene number estimate provided by genome wide analysis using
               Tetraodon nigroviridis DNA sequence
               Unpublished
               JOURNAL
               REFERENCE 2 (bases 1 to 651)
               AUTHORS Roest-Crolius,H., Jaillon,O., Dasilva,C., Fizames,C., Fisher,C.,
               Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and
               Weissenbach,J.
               TITLE Characterization and repeat analysis of the compact genome of the
               freshwater pufferfish Tetraodon nigroviridis
               JOURNAL Unpublished

```


[illegible]

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 19, 2003, 18:01:21 ; Search time 12.4727 Seconds
(without alignments)
6329.459 Million cell updates/sec

Title: US-09-997-610-1_COPY_2_156

Perfect score: 155
Sequence: 1 atagtgtctactctgtctt.....agtgaaatgacaaatgcc 155

Scoring table:
IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 424239 seqs, 254661826 residues

Total number of hits satisfying chosen parameters: 848478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published_Applications_NA:*

- 1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
- 2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
- 3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
- 4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
- 5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
- 6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
- 7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
- 8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
- 9: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
- 10: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
- 11: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
- 12: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
- 13: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
- 14: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	155	100.0	1381	9	US-09-997-610-1
2	155	100.0	1731	9	US-09-997-610-5
3	111.2	71.7	1377	9	US-09-997-610-3
4	111.2	71.7	1731	9	US-09-997-610-7
5	35.8	23.1	3477	9	US-09-935-868-25
6	35.8	23.1	3507	9	US-09-935-868-23
7	35.2	22.7	1234	9	US-09-954-531-1366
8	35	22.6	3226	10	US-09-954-456-725
9	33.4	21.5	392	10	US-09-960-352-1786
10	33.4	21.5	415	10	US-09-960-352-9137
11	33.2	21.4	549	10	US-09-923-779-49
12	33.2	21.4	561	10	US-09-923-779-47
13	33.2	21.4	617	10	US-09-923-779-51
14	33.2	21.4	648	10	US-09-923-779-66
15	33.2	21.4	653	10	US-09-923-779-53
16	33.2	21.4	655	10	US-09-923-779-68
17	33.2	21.4	656	10	US-09-923-779-69
18	33.2	21.4	658	10	US-09-923-779-71
19	33.2	21.4	659	10	US-09-923-779-52

c	20	33.2	21.4	780	10	US-09-923-779-5	Sequence 5, Appl1
	21	33.2	21.4	823	10	US-09-923-779-28	Sequence 28, Appl1
	22	33.2	21.4	886	10	US-09-923-779-4	Sequence 4, Appl1
	23	32.8	21.2	6158	10	US-09-919-497-6	Sequence 6, Appl1
	24	32.8	21.2	6158	10	US-09-954-456-762	Sequence 762, App
	25	32	20.6	990	9	US-09-738-626-2514	Sequence 2514, Ap
c	26	32	20.6	3309400	9	US-09-738-626-1	Sequence 1, Appl1
	27	31.6	20.4	1344	10	US-09-925-299-44	Sequence 44, Appl1
c	28	31.4	20.3	9388	9	US-09-924-400-141	Sequence 141, App
c	29	31.4	20.3	9388	10	US-09-810-936-141	Sequence 141, App
c	30	31.4	20.3	9388	10	US-09-429-755-141	Sequence 141, App
c	31	31.2	20.1	1619	9	US-09-764-868-400	Sequence 400, App
c	32	30.4	19.6	409	12	US-10-044-090-524	Sequence 524, App
	33	30.4	19.6	821	9	US-09-729-658B-7	Sequence 7, Appl1
	34	30.4	19.6	1176	9	US-09-729-658B-14	Sequence 14, Appl1
	35	30.4	19.6	1574	9	US-09-729-658B-1	Sequence 1, Appl1
	36	30.4	19.6	2686	9	US-09-764-868-48	Sequence 48, Appl1
	37	30.4	19.6	3380	10	US-09-799-799-1	Sequence 1, Appl1
	38	30.4	19.6	3394	10	US-09-880-107-2178	Sequence 2178, Ap
	39	30.2	19.5	162	10	US-09-783-590-11563	Sequence 11563, A
	40	30.2	19.5	88191	10	US-09-799-799-3	Sequence 3, Appl1
	41	29.6	19.1	2823	10	US-09-919-497-7	Sequence 7, Appl1
	42	29.6	19.1	32183	10	US-09-764-869-1494	Sequence 1494, Ap
c	43	29.4	19.0	1320	10	US-09-815-242-7615	Sequence 7615, Ap
	44	29.2	18.8	2379	9	US-10-063-547-123	Sequence 123, App
	45	29.2	18.8	2379	9	US-10-038-072-535	Sequence 535, App

ALIGNMENTS

RESULT 1

US-09-997-610-1
; Sequence 1, Application US/09997610
; Patent No. US20020156244A1
; GENERAL INFORMATION:
; APPLICANT: Fox, Brian
; APPLICANT: Holloway, James L.
; TITLE OF INVENTION: ADIPOCYTE COMPLIMENT RELATED PROTEIN
; TITLE OF INVENTION: ZACRP13
; FILE REFERENCE: 00-96
; CURRENT APPLICATION NUMBER: US/09/997,610
; CURRENT FILING DATE: 2001-11-29
; PRIOR APPLICATION NUMBER: US 60/253,924
; PRIOR FILING DATE: 2000-11-29
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 1381
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (2)...(1381)
US-09-997-610-1

Query Match	100.0%	Score 155;	DB 9;	Length 1381;
Best Local Similarity	100.0%;	Pred. No. 3.4e-41;		
Matches 155;	Conservative 0;	Mismatches 0;	Gaps 0;	
Qy	1	ATAGTGGTCATACCTGCTTCTTAATACGGCAGTCATTTAGCATGTAGAAAGTTGCTGGACCT	60	
Db	2	ATAGTGGTCATACCTGCTTCTTAATACGGCAGTCATTTAGCATGTAGAAAGTTGCTGGACCT	61	
Qy	61	CCAGCACACCCAGCCGCCAGAGAAAGTGGGGCTCTCTGGTGCCACCAAGTTTACCAAA	120	
Db	62	CCAGCACACCCAGCCGCCAGAGAAAGTGGGGCTCTCTGGTGCCACCAAGTTTACCAAA	121	
Qy	121	TATACAGAGAAATAGTGAATGACAAATGCC	155	
Db	122	TATACAGAGAAATAGTGAATGACAAATGCC	156	

RESULT 2
US-09-997-610-5
; Sequence 5, Application US/09997610
; Patent No. US20020156244A1
; GENERAL INFORMATION:
; APPLICANT: Fox, Brian
; APPLICANT: Holloway, James L.
; TITLE OF INVENTION: ADIPOCYTE COMPLIMENT RELATED PROTEIN
; TITLE OF INVENTION: ZACRP13
; FILE REFERENCE: 00-96
; CURRENT APPLICATION NUMBER: US/09/997,610
; CURRENT FILING DATE: 2001-11-29
; PRIOR APPLICATION NUMBER: US 60/253,924
; PRIOR FILING DATE: 2000-11-29
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 5
; LENGTH: 1731
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1731)
US-09-997-610-5

Query Match 100.0%; Score 155; DB 9; Length 1731;
Best Local Similarity 100.0%; Pred. No. 3.7e-41;
Matches 155; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATAGTGGTCATACCTGCTTATAACGGCAGTCATTGACGATGTAGAGTTCTGCGACCT 60
|||||
DB 355 ATAGTGGTCATACCTGCTTATAACGGCAGTCATTGACGATGTAGAGTTCTGCGACCT 414
|||||

QY 61 CCAGCACACCCAGGCCCCCAAGAGTGGGCCCTCTCGTGGCACCAGGTTTACCACAA 120
|||||
DB 415 CCAGCACACCCAGGCCCCCAAGAGTGGGCCCTCTCGTGGCACCAGGTTTACCACAA 474
|||||

RESULT 3
US-09-997-610-3
; Sequence 3, Application US/09997610
; Patent No. US20020156244A1
; GENERAL INFORMATION:
; APPLICANT: Fox, Brian
; APPLICANT: Holloway, James L.
; TITLE OF INVENTION: ADIPOCYTE COMPLIMENT RELATED PROTEIN
; TITLE OF INVENTION: ZACRP13
; FILE REFERENCE: 00-96
; CURRENT APPLICATION NUMBER: US/09/997,610
; CURRENT FILING DATE: 2001-11-29
; PRIOR APPLICATION NUMBER: US 60/253,924
; PRIOR FILING DATE: 2000-11-29
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 3
; LENGTH: 1377
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Degenerate polynucleotide encoding a polypeptide
; OTHER INFORMATION: of SEQ ID NO:2
; NAME/KEY: variation
; LOCATION: (1)...(1377)
; OTHER INFORMATION: Each n is independently A, T, G, or C.
; NAME/KEY: misc_feature
; LOCATION: (1)...(1377)
; OTHER INFORMATION: n = A,T,C or G
US-09-997-610-3

Query Match 100.0%; Score 155; DB 9; Length 1731;
Best Local Similarity 100.0%; Pred. No. 3.7e-41;
Matches 155; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATAGTGGTCATACCTGCTTATAACGGCAGTCATTGACGATGTAGAGTTCTGCGACCT 60
|||||
DB 355 ATAGTGGTCATACCTGCTTATAACGGCAGTCATTGACGATGTAGAGTTCTGCGACCT 414
|||||

QY 61 CCAGCACACCCAGGCCCCCAAGAGTGGGCCCTCTCGTGGCACCAGGTTTACCACAA 120
|||||
DB 415 CCAGCACACCCAGGCCCCCAAGAGTGGGCCCTCTCGTGGCACCAGGTTTACCACAA 474
|||||

Query Match 71.7%; Score 111.2; DB 9; Length 1377;
Best Local Similarity 64.5%; Pred. No. 7.3e-27;
Matches 100; Conservative 22; Mismatches 33; Indels 0; Gaps 0;

QY 1 ATAGTGGTCATACCTGCTTATAACGGCAGTCATTGACGATGTAGAGTTCTGCGACCT 60
|||||
DB 1 ATHGTNGTNATHCCNGTNYTNATHACNGCGTNGTNGTNGTNGTNGTNGTNGTNGTNGTNG 60
|||||

QY 61 CCAGCACACCCAGGCCCCCAAGAGTGGGCCCTCTCGTGGCACCAGGTTTACCACAA 120
|||||
DB 61 CCNGCNCAYCCNMGNCCNGCARGGTNGCNGCCNGCNGCNGCNGCNGCNGCNGCNGCNCAR 120
|||||

QY 121 TATACAGGAGAAATAAGTGAATGACAAATGCC 155
|||||
DB 121 TAYACNGGNGARATHWSNGARATGACNAARTGYCC 155
|||||

RESULT 4
US-09-997-610-7
; Sequence 7, Application US/09997610
; Patent No. US20020156244A1
; GENERAL INFORMATION:
; APPLICANT: Fox, Brian
; APPLICANT: Holloway, James L.
; TITLE OF INVENTION: ADIPOCYTE COMPLIMENT RELATED PROTEIN
; TITLE OF INVENTION: ZACRP13
; FILE REFERENCE: 00-96
; CURRENT APPLICATION NUMBER: US/09/997,610
; CURRENT FILING DATE: 2001-11-29
; PRIOR APPLICATION NUMBER: US 60/253,924
; PRIOR FILING DATE: 2000-11-29
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 7
; LENGTH: 1731
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Degenerate polynucleotide sequence of zacrp13/zhpl
; OTHER INFORMATION: of SEQ ID NO:6
; NAME/KEY: misc_feature
; LOCATION: (1)...(1731)
; OTHER INFORMATION: n = A,T,C or G
US-09-997-610-7

Query Match 71.7%; Score 111.2; DB 9; Length 1731;
Best Local Similarity 64.5%; Pred. No. 8.1e-27;
Matches 100; Conservative 22; Mismatches 33; Indels 0; Gaps 0;

QY 1 ATAGTGGTCATACCTGCTTATAACGGCAGTCATTGACGATGTAGAGTTCTGCGACCT 60
|||||
DB 355 ATHGTNGTNATHCCNGTNYTNATHACNGCGTNGTNGTNGTNGTNGTNGTNGTNGTNG 414
|||||

QY 61 CCAGCACACCCAGGCCCCCAAGAGTGGGCCCTCTCGTGGCACCAGGTTTACCACAA 120
|||||
DB 415 CCNGCNCAYCCNMGNCCNGCARGGTNGCNGCCNGCNGCNGCNGCNGCNGCNGCNGCNCAR 474
|||||

QY 121 TATACAGGAGAAATAAGTGAATGACAAATGCC 155
|||||
DB 475 TAYACNGGNGARATHWSNGARATGACNAARTGYCC 509
|||||

RESULT 5
US-09-935-868-25
; Sequence 25, Application US/09935868
; Patent No. US20020164690A1
; GENERAL INFORMATION:
; APPLICANT: Regeneron Pharmaceuticals, Inc
; TITLE OF INVENTION: Receptor Based Antagonists, and Methods of Making and Using
; FILE REFERENCE: REG 203D
; CURRENT APPLICATION NUMBER: US/09/935,868
; CURRENT FILING DATE: 2002-04-11
; PRIOR APPLICATION NUMBER: PCT/US99/22045


```
; NUMBER OF SEQ ID NOS: 2276
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 725
; LENGTH: 3226
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-954-456-725

Query Match          22.6%; Score 35; DB 10; Length 3226;
Best Local Similarity 61.5%; Pred. No. 0.096;
Matches 56; Conservative 0; Mismatches 35; Indels 0; Gaps 0;

Qy 26 CGCGAGTCATTGAGCATGTAGAAGTTGCTGGACCTCCAGACACCCAGGCCCCAGAGAGAG 85
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 794 CAGAAGCGCATTTGGAAGCCAGGAGCTGTGGAGCCCGCCAGGCCAGCGAGGATTCAGGAA 853
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

Qy 86 AAGTGGGGCCTCCGTGGTCACCAAGGTTTACC 116
   | | | | | | | | | | | | | | | | | |
Db 854 CAAAAGGTCCTCCCTGGGGCTCCAGGAATAGC 884

RESULT 9
US-09-960-352-1786
; Sequence 1786, Application US/09960352
; Patent No. US20020137139A1
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Tao, Nengbing
; APPLICANT: Byatt, John C.
; TITLE OF INVENTION: MATHIALAGAN, NAGAPPAN
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
; FILE REFERENCE: 16511.006/37-21(10298)C
; CURRENT APPLICATION NUMBER: US/09/960,352
; CURRENT FILING DATE: 2001-09-24
; NUMBER OF SEQ ID NOS: 15112
; SEQ ID NO 1786
; LENGTH: 392
; TYPE: DNA
; ORGANISM: Bos taurus
; OTHER INFORMATION: Clone ID: 08-LIB34-023-Q1-E1-B7
US-09-960-352-1786

Query Match          21.5%; Score 33.4; DB 10; Length 392;
Best Local Similarity 62.7%; Pred. No. 0.12;
Matches 52; Conservative 0; Mismatches 31; Indels 0; Gaps 0;

Qy 46 GAAGTTGCTGGACCTCCAGCACACCCAGGCCCCAGAGAGAGTGGGGCCTCCTCGTGCA 105
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 146 GGACCTCGGGGCCACAGGACACCCAGGCCCTCCAGGTATTAAGAGGGCCTCCAGGTATA 205

Qy 106 CCAGGTTTACCACAATATACAGG 128
   | | | | | | | | | | | | | | | | | |
Db 206 AGAGGAATACCAGGTTTGCAGG 228

RESULT 10
US-09-960-352-9137
; Sequence 9137, Application US/09960352
; Patent No. US20020137139A1
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Tao, Nengbing
; APPLICANT: Byatt, John C.
; TITLE OF INVENTION: MATHIALAGAN, NAGAPPAN
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
; FILE REFERENCE: 16511.006/37-21(10298)C
; CURRENT APPLICATION NUMBER: US/09/960,352
; CURRENT FILING DATE: 2001-09-24
; NUMBER OF SEQ ID NOS: 15112
; SEQ ID NO 9137
; LENGTH: 415
```

```
; TYPE: DNA
; ORGANISM: Bos taurus
; OTHER INFORMATION: Clone ID: 39-LIB34-018-Q1-E1-B12
US-09-960-352-9137

Query Match          21.5%; Score 33.4; DB 10; Length 415;
Best Local Similarity 62.7%; Pred. No. 0.13;
Matches 52; Conservative 0; Mismatches 31; Indels 0; Gaps 0;

Qy 46 GAAGTTGCTGGACCTCCAGCACACCCAGGCCCCAGAGAGTGGGGCCTCCTCGTGCA 105
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 156 GGACCTCGGGGCCACAGGACACCCAGGCCCTCCAGGTATTAAGAGGGCCTCCAGGTATA 215

Qy 106 CCAGGTTTACCACAATATACAGG 128
   | | | | | | | | | | | | | | | | | |
Db 216 AGAGGAATACCAGGTTTGCAGG 238

RESULT 11
US-09-923-779-49
; Sequence 49, Application US/09923779
; Patent No. US20020076721A1
; GENERAL INFORMATION:
; APPLICANT: Pyle, Ruth A.
; APPLICANT: Xu, Jiangchun
; APPLICANT: Kalos, Michael D.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF PANCREATIC CANCER
; FILE REFERENCE: 210121.553
; CURRENT APPLICATION NUMBER: US/09/923,779
; CURRENT FILING DATE: 2001-08-06
; NUMBER OF SEQ ID NOS: 155
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 49
; LENGTH: 549
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 480, 498, 523, 539
; OTHER INFORMATION: n = A,T,C or G
US-09-923-779-49

Query Match          21.4%; Score 33.2; DB 10; Length 549;
Best Local Similarity 54.0%; Pred. No. 0.17;
Matches 68; Conservative 0; Mismatches 58; Indels 0; Gaps 0;

Qy 16 GTCTTTAATAACGGCAGTCATTGAGCATGTAGAAAGTTGCTGGACCTCCAGCACACCCAGG 75
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 313 GTAGCAGTAGGAGGACTCGCAGGCTATCTCGACCAAGCTGGCCCCCAGGCCCTCCCGGT 372

Qy 76 CCCCCAGAGAAGTGGGGCCTCCTCGTGACCAAGGTTTACCACAATATACAGGAGAAATA 135
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 373 CCCCCGTGATACATCTGGTCATCTGTTCCCTCGATCTCCAGGTATCCAGGAGCCCCCT 432

Qy 136 AGTGAA 141
   | | | | |
Db 433 GGTGAA 438

RESULT 12
US-09-923-779-47
; Sequence 47, Application US/09923779
; Patent No. US20020076721A1
; GENERAL INFORMATION:
; APPLICANT: Pyle, Ruth A.
; APPLICANT: Xu, Jiangchun
; APPLICANT: Kalos, Michael D.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF PANCREATIC CANCER
; FILE REFERENCE: 210121.553
; CURRENT APPLICATION NUMBER: US/09/923,779
; CURRENT FILING DATE: 2001-08-06
```

```

; NUMBER OF SEQ ID NOS: 155
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 47
; LENGTH: 561
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 544..550
; OTHER INFORMATION: n = A,T,C or G
US-09-923-779-47

```

	Query Match	21.4%	Score 33.2	DB 10	Length 561
	Best Local Similarity	61.6%	Pred. No. 0.17		
	Matches 53	Conservative 0	Mismatches 33	Indels 0	Gaps 0
QY	53	CTGGACCTCCAGCAGCACGCCACAGGCCCGCCAGAAAGATGGGGCCCTCTGGTGACACCAAGTTT	112		
Db	404	CTGGATCTCCAGGATACCAAGGACCCCTGTGTGAACCTGGGCAAGCTGGTCTCTCAGGCC	463		
QY	113	TACCAACAATATACAGGAGAAATAAGT	138		
Db	464	CTCCAGGACCTCTGGTGCTATAGT	489		

```

RESULT 13
US-09-923-779-51
/ Sequence 51, Application US/09923779
/ Patent No US2002076721A1
/ GENERAL INFORMATION:
/ APPLICANT: Pyle, Ruth A.
/ APPLICANT: Xu, Jiangchun
/ APPLICANT: Kalos, Michael D.
/ TITLE OF INVENTION: COMPOSITIONS AND
/ TITLE OF INVENTION: AND DIAGNOSIS
/ FILE REFERENCE: 210121.553
/ CURRENT APPLICATION NUMBER: US/09/9
/ CURRENT FILING DATE: 2001-08-06
/ NUMBER OF SEQ ID NOS: 155
/ SOFTWARE: FastSeq for Windows Version
/ SEQ ID NO 51
/ LENGTH: 617
/ TYPE: DNA
/ ORGANISM: Homo sapiens
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: 581, 605
/ OTHER INFORMATION: n = A,T,C or G
US-09-923-779-51

```

	Query Match	21.4%	Score 33.2;	DB 10;	Length 617;
	Best Local Similarity	61.6%	Pred. No. 0.18;		
	Matches 53;	Conservative 0;	Mismatches 33;	Indels 0;	Gaps 0;
QY	53	CTGGACCTCCAGCACACCCAGGCCCGCCAGAAAGATGGGGCCCTCTGGTGCACACAGTTT	112		
Db	404	CTGGATCTCCAGGATACCAAGGACCCCTGTGTGAACCTGGGCAAGCTGGTCTTCAGGCC	463		
QY	113	TACCACAAATATACAGGAGAAATAAGT	138		
Db	464	CTCAGGACCTCTGGTGCTATAGGT	489		

```

RESULT 14
US-09-923-779-66
; Sequence 66, Application US/09923779
; Patent No. US2002007621A1
; GENERAL INFORMATION:
; APPLICANT: Pyle, Ruth A.
; APPLICANT: Xu, Jiangchun
; APPLICANT: Kalos, Michael D.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF PANCREATIC CANCER

```

```

; FILE REFERENCE: 210121.553
; CURRENT APPLICATION NUMBER: US/09/923,779
; CURRENT FILING DATE: 2001-08-06
; NUMBER OF SEQ ID NOS: 155
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 66
; LENGTH: 648
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 642..646
; OTHER INFORMATION: n = A,T,C or G
US-09-923-779-66

```

	Query Match	21.4%	Score 33.2	DB 10	Length 648
	Best Local Similarity	61.6%	Pred. No. 0.18		
	Matches 53	Conservative 0	Mismatches 33	Indels 0	Gaps 0
Qy	53	CTGACCTCCAGCACACCCAGGCCCCAGAGAAGCTGGGGCCCTCCTGGTGCACACAGGTT	112		
Db	404	CTGATCTCCAGGATACCAAGGACCCCTGCTGAACCTGGGCAAGCTGGTCTTCAGGCC	463		
Qy	113	TACCACAATATACAGGAGAATAAGT	138		
Db	464	CTCCAGGACCTCCCTGGTGCATATAGTT	489		

```

RESULT 15
US-09-923-779-53
; Sequence 53, Application US/09923779
; Patent No. US2002007621A1
; GENERAL INFORMATION:
; APPLICANT: Pyle, Ruth A.
; APPLICANT: Xu, Jiangchun
; APPLICANT: Kalos, Michael D.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF PANCREATIC CANCER
; FILE REFERENCE: 210121.553
; CURRENT APPLICATION NUMBER: US/09/923,779
; CURRENT FILING DATE: 2001-08-06
; NUMBER OF SEQ ID NOS: 155
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 53
; LENGTH: 653
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 517, 579, 581, 603, 649
; OTHER INFORMATION: n = A,T,C or G
; US-09-923-779-53

```

	Query Match	21.4%	Score 33.2	DB 10	Length 653		
	Best Local Similarity	61.6%	Pred. NO. 0.18				
	Matches 53	Conservative 0	Mismatches 33	Indels 0	Gaps 0		
Qy	53	CTGGA	CTCCAGC	CACACCC	CAGGGCCCC	CAGAAGATGGGGCCCTCTGGTGCACCAAGTTT	112
Db	404	CTGGATCT	CAGGATACCA	AGGACCCCT	GTGTAACTGGGCAAGCTGGTCTTCAGGCC	463	
Qy	113	TACCACA	ATATACAGG	AGAAAT	TAGT	138	
Db	464	CTCCAGG	ACCTCTCTGGT	GCCTATAGGT	489		

Search completed: February 19, 2003, 23:02:39
Job time : 20.4727 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 19, 2003, 16:09:30 ; Search time 10.2324 Seconds
(without alignments)
4645.518 Million cell updates/sec

Title: US-09-997-610-1_COPY_2_156

Perfect score: 155

Sequence: 1 atagtgttcataacctgtctt.....agtgaatgacaaaatgcc 155

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued_Patents_NA.*

- 1: /cgn2_6/ptodata/2/ina/5A_COMB.seq.*
- 2: /cgn2_6/ptodata/2/ina/5B_COMB.seq.*
- 3: /cgn2_6/ptodata/2/ina/6A_COMB.seq.*
- 4: /cgn2_6/ptodata/2/ina/6B_COMB.seq.*
- 5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq.*
- 6: /cgn2_6/ptodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	35.8	23.1	1074	2	US-08-627-151A-15
2	35.8	23.1	1404	6	Sequence 15, Appli
3	35.8	23.1	1404	6	Patent No. 5171840
4	35.8	23.1	1486	4	US-08-795-473B-3
5	35.8	23.1	1486	4	Sequence 3, Appli
6	35.8	23.1	2061	6	US-09-439-856-3
7	35.8	23.1	2061	6	Patent No. 5171840
8	35.8	23.1	3319	4	US-08-795-473B-2
9	35.8	23.1	3319	4	Sequence 2, Appli
10	33.2	21.4	1881	4	US-09-439-856-2
11	31.4	20.3	1395	4	US-09-029-348-20
12	31.4	20.3	9388	4	Sequence 88, Appli
13	31.4	20.3	9388	4	US-09-149-476-88
14	31.4	20.3	9388	4	US-08-991-789A-141
15	31.4	20.3	9388	4	Sequence 141, App
16	30.8	19.9	3294	4	US-09-062-451-141
17	30.4	19.6	821	4	US-09-598-326-141
18	30.4	19.6	1574	4	US-08-923-922A-7
19	30.4	19.6	1574	4	US-09-342-681C-14
20	29.4	19.0	2543	1	US-09-342-681C-7
21	29.4	19.0	2543	3	US-09-342-681C-1
22	29.2	18.8	3492	4	US-08-159-784-4
23	29.2	18.8	4200	1	US-08-555-669-11
24	29.2	18.8	4200	1	US-09-073-663-11
25	29.2	18.8	4200	1	US-08-923-992A-9
26	29.2	18.8	4200	5	US-08-242-932-1
27	28.8	18.6	1898	1	US-08-714-481-1
28	28.8	18.6	1898	1	US-08-494-168-1
29	28.8	18.6	1898	1	US-08-342-411A-1

28	28.4	18.3	1839	1	US-08-383-744-1	Sequence 1, Appli
29	28.4	18.3	1839	2	US-08-999-336-1	Sequence 1, Appli
30	28.4	18.3	1839	5	PCT-US96-01427-1	Sequence 1, Appli
31	28.2	18.2	390	4	US-09-134-001C-635	Sequence 635, App
32	28.2	18.2	513	4	US-09-134-001C-647	Sequence 647, App
33	28.2	18.2	585	4	US-09-134-001C-726	Sequence 726, App
34	28.2	18.2	810	4	US-09-134-001C-624	Sequence 624, App
35	28.2	18.2	1721	3	US-09-173-581-16	Sequence 16, Appl
36	28.2	18.2	1721	4	US-09-420-915-16	Sequence 16, Appl
37	28.2	18.2	3552	4	US-09-134-001C-693	Sequence 693, App
38	28	18.1	1116	4	US-08-776-971-139	Sequence 139, App
39	28	18.1	1560	2	US-08-794-795-5	Sequence 5, Appli
40	28	18.1	1560	4	US-09-249-200-5	Sequence 5, Appli
41	28	18.1	1588	6	5510466-3	Patent No. 5510466
42	28	18.1	1666	2	US-08-666-367B-3	Sequence 3, Appli
43	28	18.1	1666	4	US-09-143-438-3	Sequence 3, Appli
44	28	18.1	1703	2	US-08-794-795-1	Sequence 1, Appli
45	28	18.1	1703	4	US-09-249-200-1	Sequence 1, Appli

ALIGNMENTS

RESULT 1
US-08-627-151A-15
; Sequence 15, Application US/08627151A
; Patent No. 5866341
; GENERAL INFORMATION:
; APPLICANT: SPINELLA, Dominic
; APPLICANT: BECHERER, Kathleen
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR
; TITLE OF INVENTION: SCREENING DRUG LIBRARIES
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Gen-Probe Incorporated
; STREET: 10210 Genetic Center Drive
; CITY: San Diego
; STATE: CA
; COUNTRY: USA
; ZIP: 92121
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA: US/08/627,151A
; APPLICATION NUMBER: US/08/627,151A
; FILING DATE: 03-APR-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Fisher, Carlos A
; REGISTRATION NUMBER: 36,510
; REFERENCE/DOCKET NUMBER: CHI016
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619-410-8926
; TELEFAX: 619-410-8928
; TELEX:
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1074 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-627-151A-15

Query Match 23.1%; Score 35.8; DB 2; Length 1074;
Best Local Similarity 54.1%; Pred. No. 0.026; 62; Indels 0; Gaps 0;
Matches 73; Conservative 0; Mismatches 62;

QY 10 ATACCTGCTTTAATAACGGCAGTCATGTAGCATGTAGAGTTGCTGGACCTCCAGCACAC 69
|||||
Db 768 ATATCGGGCTGAACGGTCAAGCATGTACAAACATGGATGGTCAAGGACCTCCAGCATCA 827
QY 70 CCCAGGCCCCAGAGAAGAGTGGGGCTCTGTGTGCACCAAGTTTACCACAATATACAGGA 129
|||||
Db 828 CTGTGTCATCCACGACGCTGGAGCGGCTGAGGCACGTGGTGCACGCTTCGTGCCCAGGA 887

QY 130 GAAATAAGTGAATG 144
|||||
Db 888 GGAGTTCGGGCAAGG 902

RESULT 2
5171840-8
; PATENT NO. 5171840
; APPLICANT: KISHIMOTO, TADAMITSU
; TITLE OF INVENTION: RECEPTOR PROTEIN FOR HUMAN B CELL
; STIMULATORY FACTOR-2
; NUMBER OF SEQUENCES: 11
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/298,694
; FILING DATE: 19-JAN-1989
; SEQ ID NO: 8:
; LENGTH: 1404
5171840-8

Query Match 23.1%; Score 35.8; DB 6; Length 1404;

Best Local Similarity 54.1%; Pred. No. 0.029; Mismatches 0; Indels 0; Gaps 0;

QY 10 ATACCTGCTTTAATAACGGCAGTCATGTAGCATGTAGAGTTGCTGGACCTCCAGCACAC 69
|||||
Db 768 ATATCGGGCTGAACGGTCAAGCATGTACAAACATGGATGGTCAAGGACCTCCAGCATCA 827
QY 70 CCCAGGCCCCAGAGAAGTGGGGCTCTGTGTGCACCAAGTTTACCACAATATACAGGA 129
|||||
Db 828 CTGTGTCATCCACGACGCTGGAGCGGCTGAGGCACGTGGTGCACGCTTCGTGCCCAGGA 887
QY 130 GAAATAAGTGAATG 144
|||||
Db 888 GGAGTTCGGGCAAGG 902

RESULT 3
5480796-8
; PATENT NO. 5480796
; APPLICANT: KISHIMOTO, TADAMITSU
; TITLE OF INVENTION: ANTIBODIES AGAINST THE RECEPTOR PROTEIN
; FOR HUMAN B CELL STIMULATORY FACTOR-2
; NUMBER OF SEQUENCES: 8
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/907,650
; FILING DATE: 02-JUL-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 298,694
; FILING DATE: 19-JAN-1989
; SEQ ID NO: 8:
; LENGTH: 1404
5480796-8

Query Match 23.1%; Score 35.8; DB 6; Length 1404;

Best Local Similarity 54.1%; Pred. No. 0.029; Mismatches 0; Indels 0; Gaps 0;

QY 10 ATACCTGCTTTAATAACGGCAGTCATGTAGCATGTAGAGTTGCTGGACCTCCAGCACAC 69
|||||
Db 768 ATATCGGGCTGAACGGTCAAGCATGTACAAACATGGATGGTCAAGGACCTCCAGCATCA 827
QY 70 CCCAGGCCCCAGAGAAGTGGGGCTCTGTGTGCACCAAGTTTACCACAATATACAGGA 129
|||||
Db 828 CTGTGTCATCCACGACGCTGGAGCGGCTGAGGCACGTGGTGCACGCTTCGTGCCCAGGA 887

QY 130 GAAATAAGTGAATG 144
|||||
Db 888 GGAGTTCGGGCAAGG 902

RESULT 4
US-08-795-473B-3
; Sequence 3, Application US/08795473B
; Patent No. 6217858
; GENERAL INFORMATION:
; APPLICANT: Galun, Eithan
; APPLICANT: Nahot, Orit
; APPLICANT: Blum, Herbert E.
; TITLE OF INVENTION: A Pharmaceutical Composition for Treating
; TITLE OF INVENTION: Hepatitis B Virus (HBV) Infection
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Davidson, Davidson and Kappel, LLC
; STREET: 1140 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: MS-DOS EDITOR
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/795,473B
; FILING DATE: 11-FEB-1997
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Davidson, Clifford M.
; REGISTRATION NUMBER: 32,728
; REFERENCE/DOCKET NUMBER: 963.1007
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212)-997-1028
; TELEFAX: (212)-997-1037
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1486 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: unknown
US-08-795-473B-3

Query Match 23.1%; Score 35.8; DB 4; Length 1486;
Best Local Similarity 54.1%; Pred. No. 0.029;
Matches 73; Conservative 0; Mismatches 62; Indels 0; Gaps 0;

QY 10 ATACCTGCTTTAATAACGGCAGTCATGTAGCATGTAGAGTTGCTGGACCTCCAGCACAC 69
|||||
Db 819 ATATCGGGCTGAACGGTCAAGCATGTACAAACATGGATGGTCAAGGACCTCCAGCATCA 878
QY 70 CCCAGGCCCCAGAGAAGTGGGGCTCTGTGTGCACCAAGTTTACCACAATATACAGGA 129
|||||
Db 879 CTGTGTCATCCACGACGCTGGAGCGGCTGAGGCACGTGGTGCACGCTTCGTGCCCAGGA 938
QY 130 GAAATAAGTGAATG 144
|||||
Db 939 GGAGTTCGGGCAAGG 953

RESULT 5
US-09-439-856-3
; Sequence 3, Application US/09439856
; Patent No. 6410009
; GENERAL INFORMATION:
; APPLICANT: Galun, Eithan
; APPLICANT: Nahot, Orit
; APPLICANT: Blum, Herbert E.
; TITLE OF INVENTION: A Pharmaceutical Composition for Treating


```

RESULT 12
US-08-991-789A-141/c
; Sequence 141, Application US/08991789A
; Patent No. 6225034
; GENERAL INFORMATION:
; APPLICANT: Frudakis, Tony N.
;              Smith, John M.
;              Reed, Steven G.
; TITLE OF INVENTION: COMPOSITIONS
;              TREATMENT AND
; NUMBER OF SEQUENCES: 292
; CORRESPONDENCE ADDRESS:
; ADDRESS: Seed IP Law Group
; STREET: 701 Fifth Avenue, Suite 200
; CITY: New York, NY 10017
; COUNTRY: USA
; ATTORNEY: Seed IP Law Group
;              701 Fifth Avenue, Suite 200
;              New York, NY 10017
;              USA
; E-MAIL: seedip@seedip.com
; TEL: 212-360-0000
; FAX: 212-360-0000
;

```

```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999

```

```

;
; ATTORNEY/AGENT INFORMATION:
; NAME: Makl, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.419C2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 141:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 9388 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-09-062-451-141

Query Match 20.3%; Score 31.4; DB 4; Length 9388;
Best Local Similarity 51.8%; Pred. No. 1.5;
Matches 71; Conservative 0; Mismatches 66; Indels 0; Gaps 0;

QY 16 GTCCTATATACGGCAGTCATTGACGATGTAGAACTGCTGCACCTCCACCAATATACAGGAGAATA.135
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 8230 GCGTAAAGCTCGAAATCTCTCGGCTCAAGAGGGGCTAGATTGCTTTTATACTTTGG 8171
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 76 CCCCAGAGAAGTGGGGCCCTCTGTCGACCAAGTTTACCAACAATATACAGGAGAATA.135
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 8170 TTTAGAAAGGAGAGGGGGGTCTAGTTAAACAATTTTACAGAAATAAGTAGCAAAA 8111
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 136 AGTGAATGACAAAATG 152
   ||| ||| ||| ||| |||
Db 8110 AGTTAAAGGATAATG 8094
   ||| ||| ||| ||| |||

RESULT 14
US-09-598-326-141/c
; Sequence 141, Application US/09598326
; Patent No. 6423496
; GENERAL INFORMATION:
; APPLICANT: Fridakis, Tony N.
; Smith, John M.
; Reed, Steven G.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
; TREATMENT AND DIAGNOSIS OF BREAST CANCER
; NUMBER OF SEQUENCES: 247
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Seed Intellectual Property Law Group PLLC
; STREET: 701 Fifth Avenue, Suite 6300
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/598,326
; FILING DATE: 20-Jun-2000
; CLASSIFICATION: <unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Potter, Jane E.R.
; REGISTRATION NUMBER: 33,332
; REFERENCE/DOCKET NUMBER: 210121.419D1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 141:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 9388 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 141:

```

Db 2626 ATTCCGGGAACCCCTAAGACTCCAGACGTCCCTAACCTTCCAGACGTCCCTTAAGCTTCCA 2685

Qy 82 GAAGAAGTGGGGCTCCTGCTGGTGCACCAAGTTTACCACA 119

Db 2686 GACGTCCCTAAGCTTCCAGATGCACCGAAGTTACCACA 2723

Search completed: February 19, 2003, 22:59:00
Job time : 21.2324 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 19, 2003, 15:08:07 ; Search time 53.6445 Seconds
(without alignments)
6506.903 Million cell updates/sec

Title: US-09-997-610-1_COPY_2_156
 perfect score: 155
 Sequence: 1atagtgttcatacctgtctt.....agtgaatatgacaaatgcc 155

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N_Geneseq_101002.*

1:	/SID52/cgcdata/geneseq/geneseqn-emb1/NA1980.DAT.*
2:	/SID52/cgcdata/geneseq/geneseqn-emb1/NA1981.DAT.*
3:	/SID52/cgcdata/geneseq/geneseqn-emb1/NA1982.DAT.*
4:	/SID52/cgcdata/geneseq/geneseqn-emb1/NA1983.DAT.*
5:	/SID52/cgcdata/geneseq/geneseqn-emb1/NA1984.DAT.*
6:	/SID52/cgcdata/geneseq/geneseqn-emb1/NA1985.DAT.*
7:	/SID52/cgcdata/geneseq/geneseqn-emb1/NA1986.DAT.*
8:	/SID52/cgcdata/geneseq/geneseqn-emb1/NA1987.DAT.*
9:	/SID52/cgcdata/geneseq/geneseqn-emb1/NA1988.DAT.*
10:	/SID52/cgcdata/geneseq/geneseqn-emb1/NA1989.DAT.*
11:	/SID52/cgcdata/geneseq/geneseqn-emb1/NA1990.DAT.*
12:	/SID52/cgcdata/geneseq/geneseqn-emb1/NA1991.DAT.*
13:	/SID52/cgcdata/geneseq/geneseqn-emb1/NA1992.DAT.*
14:	/SID52/cgcdata/geneseq/geneseqn-emb1/NA1993.DAT.*
15:	/SID52/cgcdata/geneseq/geneseqn-emb1/NA1994.DAT.*
16:	/SID52/cgcdata/geneseq/geneseqn-emb1/NA1995.DAT.*
17:	/SID52/cgcdata/geneseq/geneseqn-emb1/NA1996.DAT.*
18:	/SID52/cgcdata/geneseq/geneseqn-emb1/NA1997.DAT.*
19:	/SID52/cgcdata/geneseq/geneseqn-emb1/NA1998.DAT.*
20:	/SID52/cgcdata/geneseq/geneseqn-emb1/NA1999.DAT.*
21:	/SID52/cgcdata/geneseq/geneseqn-emb1/NA2000.DAT.*
22:	/SID52/cgcdata/geneseq/geneseqn-emb1/NA2001A.DAT.*
23:	/SID52/cgcdata/geneseq/geneseqn-emb1/NA2001B.DAT.*
24:	/SID52/cgcdata/geneseq/geneseqn-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	59	38.1	1338	24 AAL440866	Human genset metab
2	38	24.5	328	21 AAC01552	Human secreted pro
3	36.8	23.7	6512	24 AB199819	Mouse ischaemic co
4	35.8	23.1	975	21 AAZ70701	Human interleukin
5	35.8	23.1	1035	21 AAZ40288	SR345 coding seque
6	35.8	23.1	1074	18 AAU04440	Interleukin 6 rece
7	35.8	23.1	1260	20 AAZ09202	Human IL-6 recepto
8	35.8	23.1	1486	19 AAV60296	Human interleukin-
9	35.8	23.1	1545	21 AAAT0763	IL-6R/IL-6 fusion

10	35.8	23.1	1627	18	AAAT97848	Human fusion polyp
11	35.8	23.1	2066	14	AAAT90340	Sequence encoding
12	35.8	23.1	2061	14	AAO41746	IL-6 receptor codi
13	35.8	23.1	2087	10	AAAT90847	DNA contg. region
14	35.8	23.1	3319	17	AAAT31441	Interleukin-6 rece
15	35.8	23.1	3319	19	AAAT60295	Human interleukin-
16	35.8	23.1	3319	21	AAAT21364	Human low adenosi
17	35.8	23.1	3319	21	AAAT35212	Human adenosi re
18	35.8	23.1	3319	24	AAAT84527	Human cDNA differe
19	35.8	23.1	3477	21	AAAT90947	Fusion polypeptide
20	35.8	23.1	3507	21	AAAT09046	Fusion polypeptide
21	35.8	23.1	4513	21	AAAT21365	Human low adenosi
22	35.8	23.1	4873	21	AAAT35243	Human adenosi re
23	35.8	22.7	1234	24	ABLT40096	Breast cancer rela
24	35.2	22.6	3226	24	ABLT84606	Human cDNA differe
25	35	22.6	3226	24	ABLT65415	Lung cancer relate
26	34.8	22.5	1121	23	AAAT89670	DNA encoding novel
27	34.8	22.5	5467	22	AAAT98343	Human EST-derived
28	34.8	22.5	5468	22	AAAT98411	Human EST-derived
29	34.6	22.3	4428	22	AAAT06578	Porcine alpha(III)
30	34.4	22.2	267	24	ABNT20477	Human ORFX polynuc
31	34.2	22.1	1518	22	AAAT01474	Human secreted pro
32	34.2	22.1	1730	21	AAAT36383	Human secreted pro
33	34.2	22.1	2592	20	AAAT41355	Human normal uteru
34	34.2	22.1	3954	21	AAAT57533	Human ORFX ORF1288
35	34.2	22.1	4428	22	AAAT06574	Bovine alpha(III)
36	34.2	22.1	4428	22	AAAT06575	Bovine alpha(III)
37	34.2	22.1	9287	24	ABKT64501	Bovine benign prost
38	33.4	21.5	433	21	AAAT61719	Clone VGT22 of a g
39	33.2	21.4	549	24	ABKT44109	cDNA #49 encoding
40	33.2	21.4	561	24	ABKT44107	cDNA #47 encoding
41	33.2	21.4	617	24	ABKT44111	cDNA #51 encoding
42	33.2	21.4	648	24	ABKT44126	cDNA #66 encoding
43	33.2	21.4	653	24	ABKT44113	cDNA #53 encoding
44	33.2	21.4	655	24	ABKT44128	cDNA #68 encoding
45	33.2	21.4	656	24	ABKT44129	cDNA #69 encoding

ALIGNMENTS

RESULT 1	
AAAL44066	
ID	AAAL44066 standard; cDNA; 1338 BP.
XX	
AC	AAAL44066;
XX	
DT	27-SEP-2002 (first entry)
XX	
DE	Human gnsset metabolic gene (GMG-9) cDNA sequence.
XX	
KW	Human; gene; ss; gene therapy; gnsset metabolic gene
KW	GMG-8; GMG-9; GMG-10; GMG-11; metabolic-related di
KW	impaired glucose tolerance; insulin resistance; Sy
KW	Type II diabetes; hyperlipidaemia; atherosclerosis
KW	heart disease; cardiac insufficiency; coronary ins
KW	high blood pressure; insulin sensitiser;
KW	non-insulin dependent diabetes mellitus.
XX	
OS	Homo sapiens.
XX	
FH	Key
FT	CDS
FT	Location/Qualifiers
FT	1..1338
FT	/*tag= a
FT	/partial
FT	/product= "Human GMG-9 protein"
FT	/note= "No stop codon is given"
XX	
PN	WO200255694-A2.
XX	
XX	
PD	18-JUL-2002.
XX	
PF	15-JAN-2002: 2002WO-IB01215.

PT genes -

XX Clalm 2; Page 2340-2352; 2690pp; English.

XX The present invention describes a method for examining ischaemic

CC conditions, comprising measuring the expression levels of particular

CC genes (I) in a test sample or determining the expression profile of a

CC gene group in the sample comprising genes selected from (I). The method

CC is useful for examining the ischaemic condition (e.g. compressive

CC ischaemia, occlusive ischaemia or vasospastic ischaemia) by measuring

CC expression levels of particular genes (ABI99202 to ABI9912, encoding

CC the protein sequences in ABB57020 to ABB57374) or by determining the

CC expression profile of a gene group comprising these genes. The

CC expression levels or expression profiles produced by these genes are

CC used as an indicator when screening for ischaemic condition-improving

CC drugs or therapeutics for ischaemic diseases. ABI9913 and ABI9914

CC represent PCR primers for a mouse ischaemic condition related sequence,

CC which are used in the exemplification of the present invention.

XX Sequence 6512 BP; 1553 A; 1798 C; 1950 G; 1211 T; 0 other;

SQ

Query Match 23.7%; Score 36.8; DB 24; Length 6512;

Best Local Similarity 61.5%; Pred. No. 0.16;

Matches 59; Conservative 0; Mismatches 37; Indels 0; Gaps 0;

QY 53 CTGGACCTCCAGCACACCCAGGCCCCAGAGAAAGTGGGGCCTCTCTGGTGCACAGGTT 112

DB 2692 CAGGACAGTCAAGGCTCCCTGGCCTCTCTGGACAGCAGGGACACCTGGAGTTCAGGGT 2751

QY 113 TACCACAATATACAGAGAGAAATAGTGAATGACAA 148

DB 2752 TCCCAGGTTCTAAAGGTGAATGGTGTGTCATGGGAA 2787

RESULT 4

AAA70701

ID AAA70701 standard; cDNA; 975 BP.

AC AAA70701;

XX

XX 11-DEC-2000 (first entry)

DE Human Interleukin 6 receptor cDNA.

XX Human; Interleukin-6 receptor; fungus; Pichia pastoris; PCR primer; ss;

KW expresslon vector; immunoglobulin-like region; cytokine receptor region.

KW Homo sapiens.

OS

XX Key Location/Qualifiers

PH 1..975

FT CDS

FT /*tag= a

FT /partial

FT /product= "human IL-6R amino acids 20-344"

FT /note= "no start or stop codon is given at the 5' or

FT 3' ends of the sequence"

FT /transl_except= (pos:169..171,aa:Glu)

FT /transl_except= (pos:841..843,aa:Gly)

XX

PN JP2000157280-A.

XX

XX 13-JUN-2000.

XX

XX 26-NOV-1998; 98JP-0335464.

XX

PR 26-NOV-1998; 98JP-0335464.

XX

PA (TOYJ) TOSOH CORP.

XX

XX WPI; 2000-468203/41.

DR P-PSDB; AAB15389.

XX

PT Yeasts transformed with IL-6 receptor gene -

XX Example 1; Page 6-8; 10pp; Japanese.

XX The invention relates to the production of human Interleukin-6 receptor

CC (IL-6R) protein in the fungus Pichia pastoris. The fungus was

CC transformed with an expression vector (pPIC9-A20LL) containing a gene

CC encoding an IL-6R protein having an immunoglobulin-like region and

CC cytokine receptor region and spanning amino acids from Leu20-Ala323. The

CC human IL-6R region was PCR amplified using the primers AAA70702-A70703.

CC This sequence represents the coding region for the IL-6R protein of the

CC invention.

XX Sequence 975 BP; 208 A; 289 C; 290 G; 188 T; 0 other;

SQ

Query Match 23.1%; Score 35.8; DB 21; Length 975;

Best Local Similarity 54.1%; Pred. No. 0.16;

Matches 73; Conservative 0; Mismatches 62; Indels 0; Gaps 0;

QY 10 ATACTGTCTTAATAACGGCAGTCATTGAGCATGTAGAAAGTTGCTGGACCTCCAGCACAC 69

DB 711 ATATCGGGCTGAACGGTCAAAAGACATTCAACACATGGTGAAGGACCTCCAGCATCA 770

QY 70 CCCAGGCCCCAGAGAAAGTGGGGCCTCTCTGGTGCACCCAGGTTTACCACAAATATACAGGA 129

DB 771 CTGTGTCTATCCACGCGCTCGAGGGGCTGAGGACGCTGTCGTCGTCGTCGTCGTCGTCG 830

QY 130 GAAATAAGTGAATG 144

DB 831 GGAGTTCGGGCAAGG 845

RESULT 5

AAZ40288

ID AAZ40288 standard; DNA; 1035 BP.

XX

AC AAZ40288;

XX

XX 25-FEB-2000 (first entry)

DT

XX SR345 coding sequence.

DE

DE Gene isolation; membrane-bound protein; fuslon protein; drug production;

KW antigen-binding cell; secretable functional protein; antigenic protein;

KW protein isolation; diagnosis; SR345 protein; ss.

XX

OS Homo sapiens.

XX

PN WO9960113-A1.

XX

XX 25-NOV-1999.

XX

XX 30-APR-1999; 99WO-JP02341.

XX

PR 20-MAY-1998; 98JP-0138652.

PR 01-OCT-1998; 98JP-0279876.

XX

PA (CHUS) CHUGAI SEIYAKU KK.

XX

PI Tsuchiya M, Salto M, Ohtomo T;

XX

DR WPI; 2000-039382/03.

DR P-PSDB; AAY55071.

XX

XX Efficient and selective isolation of a gene encoding membrane protein

PT with low or no antigenic binding activity, for diagnosis, study of, and

PT production of drugs treating abnormal functions of the protein -

XX

XX Example 1; Page 49-52; 120pp; Japanese.

XX This sequence encodes the SR345 protein.

CC The invention relates to a method for isolating a gene encoding a

CC membrane-bound protein, comprising introducing a vector into a cell,

CC contacting an antigen with the cell expressing the fused protein encoded

CC cells and platelets. Transmission of a signal of IL-6 to target cells for
CC stimulation with reduced antigenicity is possible. This sequence encodes
CC the IL-6 receptor/IL-6 fusion protein described in the invention.
XX
SQ Sequence 1260 BP; 341 A; 348 C; 321 G; 250 T; 0 other;

Query Match 23.1%; Score 35.8; DB 20; Length 1260;
Best Local Similarity 54.1%; Pred. No. 0.17;
Matches 73; Conservative 0; Mismatches 62; Indels 0; Gaps 0;
QY 10 ATACCTGCTTAAATACGGCAGTCATTGAGCATGTAGAAAGTTGCTGGACCTCCAGCACAC 69
DB 423 ATATCGGGCTGAACGGTCAAGACATTCACAACTGGAATGGTCAAGGACCTCCAGCATCA 482
QY 70 CCCAGGCCCCAGAGAAGTGGGGCTCTGCTGTGCACCAAGTTTACCACAAATATACAGGA 129
DB 483 CTGTGTCATCCAGCAGCGCTGGAGGGCTGAGGCACGTGTGCAGCTCTGTGCCCCAGGA 542
QY 130 GAAATAAGTGAANTG 144
DB 543 GGAGTTCGGGCAAGG 557

RESULT 8
AAV60296
ID AAV60296 standard; DNA; 1486 BP.
XX
AC AAV60296;
XX
DT 02-FEB-1999 (first entry)
XX
DE Human interleukin-6 nucleotide sequence.
XX
KW Interleukin-6; human; hepatitis B virus; HBV; infection; therapy;
KW ss.
XX
OS Homo sapiens.
XX
PN WO9835694-A2.
XX
PD 20-AUG-1998.
XX
PF 10-FEB-1998; 98WO-US08898.
XX
PR 11-FEB-1997; 97US-0795473.
XX
PA (DAVI/) DAVIDSON C M.
PA (HADA-) HADASIT MEDICAL RES SERVICES & DEV.
XX
WPI; 1998-520755/44.
XX
PT Treatment of hepatitis B virus infection - using a soluble active
PT agent which prevents interaction of HBV with hepatocytes mediated by
PT human interleukin 6
XX
PS Disclosure; Fig 7; 51pp; English.
XX

This is a previously reported nucleotide sequence for human
interleukin-6 (hIL-6). The invention relates to the finding that
hIL-6 is essential for hepatitis B virus (HBV) infection. The
invention provides a pharmaceutical composition for the treatment
of HBV infection, comprising a soluble active agent that interacts
with at least one of the binding sites between hIL-6 and the pSI
region of HBV and between hIL-6 and hepatocytes and other
HBV-permissive cells. The active agent competitively binds to at
least one of these sites and thereby prevents hIL-6-mediated HBV
infection of hepatocytes and other HBV-permissive cells. The
soluble active agent is selected from glycoprotein 80 (gp80) having
receptor sites which interact with hIL-6, soluble glycoprotein 130
(gp130) having receptor sites which interact with hIL-6, hIL-6
derived peptide Lys41-Ala56, hIL-6 derived peptide Gly77-Glu95,
hIL-6 derived peptide Gln153-His165, a combined 1 and 2 hIL-6
mutant (mhIL-6 1+2), and mhIL-6 1+2 substituted with Phe171 to Leu

CC and Ser177 to Arg, and mixtures of any of these.

XX Sequence 1486 BP; 305 A; 453 C; 439 G; 289 T; 0 other;

Query Match 23.1%; Score 35.8; DB 19; Length 1486;
Best Local Similarity 54.1%; Pred. No. 0.18;
Matches 73; Conservative 0; Mismatches 62; Indels 0; Gaps 0;
QY 10 ATACCTGCTTAAATACGGCAGTCATTGAGCATGTAGAAAGTTGCTGGACCTCCAGCACAC 69
DB 819 ATATCGGGCTGAACGGTCAAGACATTCACAACTGGAATGGTCAAGGACCTCCAGCATCA 878
QY 70 CCCAGGCCCCAGAGAAGTGGGGCTCTGCTGTGCACCAAGTTTACCACAAATATACAGGA 129
DB 879 CTGTGTCATCCAGCAGCGCTGGAGGGCTGAGGCACGTGTGCAGCTCTGTGCCCCAGGA 938
QY 130 GAAATAAGTGAANTG 144
DB 939 GGAGTTCGGGCAAGG 953

RESULT 9
AAA70763
ID AAA70763 standard; DNA; 1545 BP.
XX
AC AAA70763;
XX
DT 17-JAN-2001 (first entry)
XX
DE IL-6R/IL-6 fusion protein coding sequence.
XX
KW Fusion protein; human; interleukin 6 receptor; Pichia pastoris; yeast;
KW drug; myeloid stem cell; platelet; blood; ss.
XX
OS Homo sapiens.
OS Synthetic.
XX
FH Key Location/Qualifiers
FT CDS 1..1545
FT /*tag= a
FT /product= "IL-6R/IL-6 fusion protein"
FT /transl_except= (pos:169..171,aa:Glu)
FT /transl_except= (pos:841..843,aa:Gly)
XX
PN JP2000166539-A.
XX
PD 20-JUN-2000.
XX
PF 03-DEC-1998; 98JP-0343933.
XX
PR 03-DEC-1998; 98JP-0343933.
XX
PA (TOYJ) TOSOH CORP.
XX
WPI; 2000-485548/43.
DR P-PSDB; AABL5404.
XX
PT Yeast of Pichia Pastoris genus transformed by expression vector
PT containing gene encoding fused protein of interleukin-6 receptor
PT (IL-6R) protein for amplifying myeloid stem cells and increasing
platelets -
XX
PS Example 1; Page 8-10; 11pp; Japanese.
XX
CC The invention relates to the production of a fusion protein comprising
CC the human interleukin 6 receptor (IL-6R) fused to the human IL-6 protein
CC in a Pichia pastoris yeast cell. This sequence represents the coding
CC sequence for the IL-6R/IL-6 fusion protein which is cloned into an
CC expression vector for introduction into P. pastoris. The IL-6R protein
CC is used as a drug for amplifying myeloid stem cells and increasing
CC platelets in blood.
XX
SQ Sequence 1545 BP; 391 A; 433 C; 424 G; 297 T; 0 other;

Query Match 23.1%; Score 35.8; DB 21; Length 1545;
Best Local Similarity 54.1%; Pred. No. 0.19;
Matches 73; Conservative 0; Mismatches 62; Indels 0; Gaps 0;

QY 10 ATACCTGTCTTATAACGGCAGTCATTGACCATGTAGAAAGTTGCTGGACCTCCAGCACAC 69
|||||
Db 711 ATATCGGGCTGAACGGTCAAGACATTCACACATGGATGGTCAAGGACCTCCAGCATCA 770
|||||
QY 70 CCAGGCCCCCAAGAAAGTGGGGCCTCTCTGGTGCACCCAGGTTTACCAACAATATACAGGA 129
|||||
Db 771 CTGTGTTCATCCAGCAGCGCTGGAGCGCCTGAGGCACGTGGTGCACCTTCGTGCCCGAGGA 830
|||||
QY 130 GAAATAAGTGAATG 144
|||||
Db 831 GGAGTTCGGGCAAGG 845
|||||

RESULT 10
AAT97848
ID AAT97848 standard; DNA; 1627 BP.
XX
AC AAT97848;
XX
DT 25-MAR-1998 (first entry)
XX
DE Human fusion polypeptide H-IL-6 DNA containing a 18 amino acid linker.
XX
KW Interleukin-6; IL-6; interleukin-6 receptor; IL-6R; ligand; conjugate;
KW protein interaction; therapeutic; antagonist; ss.
XX
OS Synthetic.
OS Homo sapiens
XX
FH Key Location/Qualifiers
FT CDS 34..1611
FT sig_peptide /tag= a
FT 34..90
FT /tag= b
FT mat_peptide 91..1608
FT /tag= c
FT /product= H-IL-6
FT /note= "fusion polypeptide"

XX
PN WO9732891-A2.
XX
PD 12-SEP-1997.
XX
PE 07-MAR-1997; 97WO-DE00458.
XX
PR 07-MAR-1996; 96DE-4008813.
XX
PA (ANGE-) ANGEWANDTE GENTECHNOLOGIE SYSTEME GMBH.
XX
PI Rose-John S;
XX
WPI; 1997-470536/43.
DR P-PSDB; AAN36846.
XX
PT Conjugate of two peptide(s) with mutual affinity connected by a
PT linker - used to modulate interactions between proteins, e.g. for ex
PT vivo expansion of human stem cells
XX
PS Claim 12; Fig 1; 19pp; German.
XX
CC This sequence encodes the fusion polypeptide H-IL-6 which contains an
CC 18 amino acid linker which joins the carboxy terminus of human
CC interleukin-6 receptor (IL-6R) with the amino terminus of human
CC interleukin-6 (IL-6). Such conjugates could be used to modulate
CC interactions between proteins, particularly to overcome interrupted
CC interactions caused by an incomplete interleukin-6 (IL-6) receptor. These
CC constructs derived from IL-6 and its receptor, can also be used for ex
CC vivo expansion of human stem cells, and as a therapeutic IL-6 receptor

CC antagonist.
XX
SQ Sequence 1627 BP; 388 A; 450 C; 481 G; 308 T; 0 other;
Query Match 23.1%; Score 35.8; DB 18; Length 1627;
Best Local Similarity 54.1%; Pred. No. 0.19; Mismatches 62; Indels 0; Gaps 0;
Matches 73; Conservative 0; Mismatches 62; Indels 0; Gaps 0;

QY 10 ATACCTGTCTTATAACGGCAGTCATTGACCATGTAGAAAGTTGCTGGACCTCCAGCACAC 69
|||||
Db 801 ATATCGGGCTGAACGGTCAAGACATTCACACATGGATGGTCAAGGACCTCCAGCATCA 860
|||||
QY 70 CCAGGCCCCCAAGAAAGTGGGGCCTCTCTGGTGCACCCAGGTTTACCAACAATATACAGGA 129
|||||
Db 861 CTGTGTTCATCCAGCAGCGCTGGAGCGCCTGAGGCACGTGGTGCACCTTCGTGCCCGAGGA 920
|||||
QY 130 GAAATAAGTGAATG 144
|||||
Db 921 GGAGTTCGGGCAAGG 935
|||||

RESULT 11
AAN90340
ID AAN90340 standard; cDNA; 2061 BP.
XX
AC AAN90340;
XX
DT 31-MAR-1992 (first entry)
XX
DE Sequence encoding a receptor protein for human B cell stimulating
DE factor-2 (BSF2 receptor).
XX
KW B cell; immune disorder; therapy; diagnosis; prophylaxis; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 247..1753
FT /tag= a
FT /note= "SQ claimed"
XX
PN EP325474-A.
XX
PD 26-JUL-1989.
XX
PE 20-JAN-1989; 89EP-0300536.
XX
PR 14-JAN-1989; 89JP-0017461.
PR 22-JAN-1988; 88JP-0012387.
PR 25-JAN-1988; 88JP-0012599.
PR 04-AUG-1988; 88JP-0194885.
PR 20-JAN-1989; 89JP-0009774.
XX
PA (KISH/) KISHIMOTO T.
XX
PI Kishimoto T;
XX
WPI; 1989-214667/30.
DR P-PSDB; AAN90284.
XX
PT Receptor protein for human B cell stimulating factor-2 - obt'd. by
PT recombinant DNA techniques and used as diagnostic, prophylactic or
PT therapeutic agent
XX
PS Disclosure; Fig 3; 63pp; English.
XX
CC The cDNA in AAN90340 was derived from monocytic cell line U937.
CC Isolated BSF2 receptor and DNA encoding it are claimed, as are
CC (b) expression vectors; (c) host organisms; (d) antibodies; and
CC (e) hybridomas.
XX
SQ Sequence 2061 BP; 418 A; 631 C; 621 G; 391 T; 0 other;

Db	1074	CTGTGTATCCACGACGCGCTGGAGGCGCTGAGGCACGTGTGCAGCTTCGTGCCACGGA	1133
Qy	130	GAATAAGTGAATG	144
Db	1134	GGAGTTCGGCAAGG	1148
RESULT 14			
ID	AAT31441	standard; cDNA; 3319 BP.	
AC	AAT31441:		
XX	28-NOV-1996	(first entry)	
XX	Interleukin-6	receptor coding sequence.	
XX	Interleukin-6; IL; receptor; antisense oligonucleotide; inhibition;		
KW	gene expression; kidney tumour; myeloma; Kaposi's sarcoma; psoriasis;		
KW	rheumatoid arthritis; endotoxic shock; ss.		
XX	Homo sapiens.		
XX	Key	Location/Qualifiers	
FH	CDS	438..1844	
FT	FT	/*tag= a	
FT	FT	/product= Interleukin-6 receptor.	
XX	W09618416-AL.		
XX	20-JUN-1996.		
XX	15-DEC-1995;	95WO-JP02587.	
XX	18-AUG-1995;	95JP-0210739.	
PR	16-DEC-1994;	94JP-0313167.	
XX	(CHUS)	CHUGAI SEIYAKU KK.	
XX	Koishibara Y,	Kuromaru K;	
XX	WPI; 1996-300392/30.		
DR	P-PSDB; AAR98364.		
XX	Anti-sense oligo:nucleotide inhibitor against human IL-6R expression		
PT	- for treatment of e.g. tumours, cancers, rheumatoid arthritis,		
PT	psoriasis, endotoxic shock, etc.		
XX	Claim 2; Page 17-21; 32pp; Japanese.		
XX	Antisense oligonucleotides may be used to inhibit the expression of		
CC	the interleukin-6 receptor. Inhibition of expression of the		
CC	IL-6 receptor is useful in the treatment of kidney tumours, myeloma,		
CC	Kaposi's sarcoma, rheumatoid arthritis, psoriasis and endotoxic		
CC	shock. The antisense oligonucleotides are administered at a dosage		
CC	of 0.1-100mg/kg, pref. 0.1-50 mg/kg.		
XX	Sequence 3319 BP; 735 A; 937 C; 960 G; 587 T; 0 other;		
SQ	Query Match	23.1%; Score 35.8; DB 17; Length 3319;	
	Best Local Similarity	54.1%; Pred. No. 0.25;	
	Matches	73; Conservative 0; Mismatches 62; Indels 0; Gaps	
Qy	10	ATACCTGCTCTTAATACGGCAGTCATTTAGAGCATGTAGAAAGTTGTGGACCTCCAGCAC	69
Db	1205	ATATCGGGCTGAACGGTCAAGACATTCACAACATGGATGGTCAAGGACCTCCAGCATCA	1264
Qy	70	CCAGGCCCCACAGAAGTGGGGCTCTCTGTGCACCGAGTTTACCACATATACAGGA	129
Db	1265	CTGTGTATCCACGACGCGCTGAGGCGCGCTGAGGCACGTGTGCAGCTTCGTGCCACGGA	1324
Qy	130	GAATAAGTGAATG	144

[illegible]

Search completed: February 19, 2003, 22:56:26
Job time : 61.6445 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 19, 2003, 23:16:26 ; Search time 494.184 Seconds
(without alignments)
9128.055 Million cell updates/sec

Title: US-09-997-610-1_COPY_2_156
Perfect score: 155
Sequence: 1 atagtgtctactcttctt.....agtgaatgacaaatgcc 155

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:**

- 1: gb_ba:**
- 2: gb_htg:**
- 3: gb_in:**
- 4: gb_om:**
- 5: gb_ov:**
- 6: gb_pat:**
- 7: gb_ph:**
- 8: gb_pl:**
- 9: gb_pr:**
- 10: gb_ro:**
- 11: gb_sts:**
- 12: gb_sy:**
- 13: gb_un:**
- 14: gb_vl:**
- 15: em_ba:**
- 16: em_fun:**
- 17: em_hum:**
- 18: em_in:**
- 19: em_mu:**
- 20: em_om:**
- 21: em_or:**
- 22: em_ov:**
- 23: em_pat:**
- 24: em_ph:**
- 25: em_pl:**
- 26: em_ro:**
- 27: em_sts:**
- 28: em_un:**
- 29: em_vl:**
- 30: em_htg_hum:**
- 31: em_htg_inv:**
- 32: em_htg_other:**
- 33: em_htg_mus:**
- 34: em_htg_pln:**
- 35: em_htg_rod:**
- 36: em_htg_mam:**
- 37: em_htg_vrt:**
- 38: em_sy:**
- 39: em_htgo_hum:**
- 40: em_htgo_mus:**
- 41: em_htgo_other:**

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	111	71.6	145880	9	HS302D9	Z82198 Human DNA s
2	40.6	26.2	5851	3	AF282902	AF282902 Hydra vul
3	39.2	25.3	87100	9	HS43814	Z97635 Human DNA s
4	38.2	24.6	1395	5	DRE318214	AJ318214 Danio rer
5	38	24.5	287	3	OOS420010	AJ420010 Osteragi
6	37.8	24.4	9448	3	CEU22327	U22327 Caenorhabd
7	37.8	24.4	9541	3	CECOLA21V	Z22964 C.elegans a
8	37.8	24.4	34671	3	U53342	U53342 Caenorhabd
9	37.2	24.0	123789	2	AC110674	AC110674 Tetraodon
10	36.8	23.7	3612	10	MNCOLA4	X67777 Mouse mRNA
11	36.8	23.7	6512	6	AX306181	AX306181 Sequence
12	36.8	23.7	6512	10	MUSCOLIA4A	J04694 Mus muscu
13	36.8	23.7	7765	10	AF169387	AF169387 Mus muscu
14	36.6	23.6	694	3	HSNCOL1	X61045 Hydra N-COL
15	36.6	23.6	95492	2	AC096135	AC096135 Rattus no
16	36.2	23.4	36532	3	CEF57B1	Z78064 Caenorhabd
17	36.2	23.4	269619	3	CEY51H4A	AL132952 Caenorhab
18	35.8	23.1	975	6	E64808	E64808 yeast trans
19	35.8	23.1	1074	6	AR031384	AR031384 Sequence
20	35.8	23.1	1074	6	BD009752	BD009752 Compositi
21	35.8	23.1	1260	6	E28089	E28089 Novel IL-6
22	35.8	23.1	1486	6	ARI45506	ARI45506 Sequence
23	35.8	23.1	1486	9	HS16REC	X58298 Human mRNA
24	35.8	23.1	1545	6	E35612	E35612 IL-6 recept
25	35.8	23.1	1612	6	A93715	A93715 Sequence 2
26	35.8	23.1	1627	6	A93714	A93714 Sequence 1
27	35.8	23.1	2066	6	E04823	E04823 cDNA encodi
28	35.8	23.1	2087	6	E02673	E02673 cDNA encodi
29	35.8	23.1	3319	6	ARI45505	ARI45505 Sequence
30	35.8	23.1	3319	6	AX399205	AX399205 Sequence
31	35.8	23.1	3319	6	EL2979	EL2979 cDNA encodi
32	35.8	23.1	3319	9	HSIL6R	X12830 Human mRNA
33	35.6	23.0	3734	14	HVSX99519	X99519 Herpesvirus
34	35.6	23.0	11942	1	AE010554	AE010554 Fusobacte
35	35.4	22.8	130619	2	AC120069	AC120069 Rattus no
36	35.2	22.7	1234	6	AX331924	AX331924 Sequence
37	35.2	22.7	1234	9	HSU52521	U52521 Human arfap
38	35.2	22.7	1993	3	BMCOLGMR	Z30348 B.mori mRNA
39	35	22.6	1973	9	HS10ALCOL	X65120 H.sapiens C
40	35	22.6	1976	10	RNJ005395	AJ005395 Rattus no
41	35	22.6	2217	10	RNPRO1C	X70369 R.norvegicu
42	35	22.6	3215	9	HSCOLX3	X72580 Homo sapien
43	35	22.6	3226	6	AX333243	AX333243 Sequence
44	35	22.6	3226	9	HSCOLA1X	X60382 H.sapiens C
45	35	22.6	3226	11	G28608	G28608 human STS S

ALIGNMENTS

RESULT 1
HS302D9
LOCUS
DEFINITION Human DNA sequence from clone RP1-302D9 on chromosome 22 Contains
ACCESSION Z82198
VERSION Z82198.2 GI:6572207
KEYWORDS HTG.
SOURCE Homo sapiens
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
TITLE 1 (bases 1 to 145880)
Bridgeman,A.
Direct Submission

JOURNAL

Submitted (08-DEC-1999) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone requests: clonequests@sanger.ac.uk

On Dec 13, 1999 this sequence version replaced gi:3164067. During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

This sequence has been finished according to sequence map criteria as follows. An attempt is made to resolve all sequencing problems, such as compressions and repeats, but not necessarily within known annotated human repeat sequence elements (e.g. Alu). Where the sequence is ambiguous, there is an annotation using the 'unsure' feature key.

The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em.; EMBL; Sw.; SWISSPROT; Tr.; TREMBL; Wp.; WORMPEP; Information on the WORMPEP database can be found at http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence was generated from part of bacterial clone contigs of human chromosome 22, constructed by the Sanger Centre Chromosome 22 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/Chr22>

RP1-302D9 is from the library RP1-1 constructed at the Roswell Park Cancer Institute by the group of Pieter de Jong. For further details see <http://bacpac.med.buffalo.edu/VECTOR:PCYPAC2>

This sequence is the entire insert of clone RP1-302D9 The true left end of clone CTA-282F2 is at 69682 in this sequence. The true right end of clone CTA-415G2 is at 55167 in this sequence.

FEATURES

source

1..145980
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="22"
/clone="RP1-302D9"
/clone.lib="RP1-1"

repeat_region
188..245
/note="MER3 repeat: matches 144..209 of consensus"
246..571
/note="AluX repeat: matches 1..312 of consensus"
repeat_region
572..759
/note="MER3 repeat: matches 1..144 of consensus"
783..933
/note="MERA repeat: matches 26..187 of consensus"
repeat_region
1033..1336
/note="AluSp repeat: matches 1..299 of consensus"
repeat_region
1450..1583
/note="MIR repeat: matches 24..160 of consensus"
repeat_region
1687..1752
/note="L2 repeat: matches 2593..2661 of consensus"
repeat_region
2350..2660
/note="AluSc repeat: matches 3..309 of consensus"
repeat_region
2684..2981
/note="AluSq repeat: matches 2..300 of consensus"
repeat_region
3323..3343
/note="MLTIE repeat: matches 116..136 of consensus"
repeat_region
3344..3652
/note="AluY repeat: matches 1..309 of consensus"
repeat_region
3653..3928
/note="MLTIE repeat: matches 136..359 of consensus"
repeat_region
3929..4278
/note="THEIB repeat: matches 3..364 of consensus"
repeat_region
4279..4485
/note="MLTIE repeat: matches 359..568 of consensus"
repeat_region
5073..5176
/note="52 copies 2 mer ct 78 conserved"
repeat_region
5181..5491
/note="AluJb repeat: matches 1..311 of consensus"
repeat_region
6369..6485
/note="L2 repeat: matches 2579..2705 of consensus"

repeat_region
6647..6685
/note="WADE1 repeat: matches 1..23 of consensus"
repeat_region
6686..6987
/note="AluX repeat: matches 1..302 of consensus"
repeat_region
6988..7036
/note="WADE1 repeat: matches 23..77 of consensus"
repeat_region
7482..7754
/note="AluJb repeat: matches 9..290 of consensus"
repeat_region
7775..8060
/note="AluJo repeat: matches 1..295 of consensus"
repeat_region
8414..8551
/note="L2 repeat: matches 2553..2706 of consensus"
repeat_region
8914..9030
/note="MIR repeat: matches 147..262 of consensus"
repeat_region
9110..9280
/note="MIR repeat: matches 91..262 of consensus"
repeat_region
9283..9412
/note="MIR repeat: matches 15..144 of consensus"
repeat_region
9521..9679
/note="FAM repeat: matches 3..161 of consensus"
repeat_region
9820..10225
/note="MSTB repeat: matches 2..425 of consensus"
misc_feature
complement(10179..10678)
/note="match: GSS: Em:B56592"
misc_feature
complement(10204..10728)
/note="match: GSS: Em:AQ701486"
misc_feature
complement(10249..10706)
/note="match: GSS: Em:AQ225495"
repeat_region
10312..10383
/note="MIR repeat: matches 79..150 of consensus"
misc_feature
10718..11310
/note="match: GSS: Em:B14024"
misc_feature
10784..11201
/note="match: GSS: Em:B43656"
repeat_region
11838..11946
/note="MIR repeat: matches 20..137 of consensus"
repeat_region
12174..12445
/note="L2 repeat: matches 1988..2275 of consensus"
repeat_region
12444..12642
/note="MIR repeat: matches 63..241 of consensus"
misc_feature
13017..13369
/note="match: STS: Em:G49301"
repeat_region
13331..13397
/note="MIR repeat: matches 174..244 of consensus"
repeat_region
13398..13698
/note="AluSp repeat: matches 1..302 of consensus"
repeat_region
13699..13810
/note="MIR repeat: matches 76..174 of consensus"
repeat_region
13806..13919
/note="MIR repeat: matches 77..189 of consensus"
repeat_region
13945..14060
/note="MIR repeat: matches 24..142 of consensus"
repeat_region
14061..14367
/note="AluY repeat: matches 1..301 of consensus"
repeat_region
14368..14452
/note="MIR repeat: matches 141..225 of consensus"
repeat_region
14589..14679
/note="MIR repeat: matches 173..262 of consensus"
misc_feature
14597..15201
/note="match: GSS: Em:AQ553482"
misc_feature
14616..15060
/note="match: GSS: Em:AQ370601"
repeat_region
14868..15040
/note="MIR repeat: matches 49..233 of consensus"
repeat_region
15071..15188
/note="L2 repeat: matches 2112..2239 of consensus"
repeat_region
15304..15399
/note="MLTIB repeat: matches 1..99 of consensus"
repeat_region
15490..15662
/note="AluSg1 repeat: matches 2..114 of consensus"
repeat_region
15669..15727
/note="MLTIB repeat: matches 119..178 of consensus"
repeat_region
15728..16027

repeat_region /note="Alusc repeat: matches 1. .299 of consensus" 16028. .16245
repeat_region /note="MLR1B repeat: matches 178. .330 of consensus" 16546. .16854
repeat_region /note="Aluy repeat: matches 1. .300 of consensus" 18296. .18323
repeat_region /note="MSTA repeat: matches 2. .29 of consensus" 18324. .18392
repeat_region /note="MER66-internal repeat: matches 4919. .4993 of consensus" 18393. .18712
repeat_region /note="AlujB repeat: matches 1. .311 of consensus" 18713. .19133
repeat_region /note="MER66-internal repeat: matches 4548. .4919 of consensus" 19251. .19719
misc_feature /note="match: GSS: Em:AQ005063" 19537. .20290
misc_feature /note="match: GSS: Em:B14179" 20317. .20382
repeat_region /note="HERVH21 repeat: matches 4657. .5784 of consensus" 20513. .20666
repeat_region /note="33 copies 2 mer ta 68 conserved" 20682. .21008
repeat_region /note="77 copies 2 mer tt 70 conserved" 21239. .21553
repeat_region /note="Alusg1 repeat: matches 1. .306 of consensus" 21882. .22254
repeat_region /note="HUPERS-P3 repeat: matches 4410. .4713 of consensus" 22302. .22537
repeat_region /note="THE1B repeat: matches 1. .364 of consensus" 23905. .23989
repeat_region /note="MER66-internal repeat: matches 2186. .2417 of consensus" 22538. .22850
repeat_region /note="Alusp repeat: matches 1. .313 of consensus" 22851. .23801
repeat_region /note="MER66-internal repeat: matches 1210. .2186 of consensus" 23905. .23989
repeat_region /note="MER66-internal repeat: matches 3017. .3102 of consensus" 71.0%; Score 111; DB 9; Length 145880;
Query Match Best Local Similarity 100.0%; Pred. No. 3e-23;
Matches 111; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATAGTGGTCATACCTGCTTTAATAACGCATCATTTAGCATGTAGAGTTGCTGGACCT 60
Db 34786 ATAGTGGTCATACCTGCTTTAATAACGCATCATTTAGCATGTAGAGTTGCTGGACCT 34845
Qy 61 CCAGCACACCCAGGCCGCCAGAGAGAGTGGGGCTCTCTGTCGACACAGGT 111
Db 34846 CCAGCACACCCAGGCCGCCAGAGAGAGTGGGGCTCTCTGTCGACACAGGT 34896

RESULT 2
AF282902 5851 bp mRNA linear INV 17-DEC-2000
LOCUS Hydra vulgaris type IV collagen alpha 1 chain precursor, mRNA, complete cds.
DEFINITION AF282902
VERSION AF282902
KEYWORDS AF282902.1 GI:11875611
SOURCE Hydra vulgaris.
ORGANISM Hydra vulgaris.
Eukaryota; Metazoa; Chnidaria; Hydrozoa; Hydroida; Anthomedusae; Hydridae; Hydra.
REFERENCE 1 (bases 1 to 5851)
AUTHORS Fowler,S.J., Jose,S., Zhang,X., Deutzmann,R., Sarraz,M.P. Jr. and Boot-Handford,R.P.
TITLE Characterization of hydra type IV collagen. Type IV collagen is essential for head regeneration and its expression is up-regulated upon exposure to glucose

J. Biol. Chem. 275 (50), 39589-39599 (2000)
MEDLINE 20564332
PUBMED 10956657
REFERENCE 2 (bases 1 to 5851)
AUTHORS Fowler,S.J. and Boot-Handford,R.P.
TITLE Direct Submission
JOURNAL Submitted (27-JUN-2000) School of Biological Sciences, University of Manchester, Oxford Road, Manchester M13 9PT, UK
FEATURES
source Location/Qualifiers
1. 5851
/organism="Hydra vulgaris"
/db_xref="taxon:6087"
157. .5328
/product="type IV collagen alpha 1 chain precursor"
/protein_id="AAG40729.1"
/db_xref="GI:11875612"
/translation="MNHTTKWTIIVIFLTHSLCVFTQLCGGGCCNQCINVCVQKQ
DRGSVGLPGFKPIGERGPGGPGAGRGKGDGAGPLQAGVGARGKVGPPGPTGI
PGSHRRPGDEGRQCEPGLPCNGTKGDI GPPGPRGKDVQVQPSGLPGIGRPPGDS
TATLKQKQEPGPKSODGAPDPTKDGKESKAGRMGPQGLRGKAGRGDSNITI
FERGDKGIDILPGPRGDCNCSSTGELTINIOGPKGEOKKGDQOKGEPGPGOA
GEAQDCQKGEKDKGEISGNPSIOGEGKDGATGATGATGATGATGATGATGATGAT
DHGTGKSEGPVGDGKQESITGPPQIGERQKQGEKKGEPGEPGPPNGEVGDT
PPGLKGLKEIGMTGSGEPGPGTGAEGMKGPIGAGRGETGSKGESGRPGVQGS
PGMDNPGQIGPPGEPGIPGPPGPNPQVNDLGETVLPGPDDGQPGQIGVAGP
QSGPIRGDKDACKSCPQSGQGVAGDGLPCANGKDKDGKDEGKPGLEGSVGP
GESCLTQPGPIGEGKGTGAKNGKGRSDRIVOGERGEKTKDGMPSGSGIFGKEQ
GVFPGTDGFPQGGIQQGDRGAFKSGIIPNNGEKEKKGKNGKGLRGRGERGR
DGEPPGSGQPMOKPMPKPEPFTALGQGGDKDGVSESGPPGQAGAGVGNQGFK
IGEPGLGAPSGEGERGADGTPGAGNGENDGDDGPKGDKGKGLQPIGISGP
IKGEPGLGDPGKPLDAPKSVKPGKGTGKGTGKGTGKGTGKGTGKGTGKGTGKGT
IGPQGPDKGQDQSVGPKGTGKGTGKGTGKGTGKGTGKGTGKGTGKGTGKGTGKGT
SVGLPGDKGKSPGKGTGKGTGKGTGKGTGKGTGKGTGKGTGKGTGKGTGKGTGKGT
EKVGTGTLGDKGEPGYSQKAKGTGVRGKPGATSNVMGPKNGRSGESAKGEPG
KQPPGAGGQGEGERGIGTDKGTGDTGDIQGPAGGERGERGERGERGERGERGER
GTLKGNPGEPRGPTGPPGAGKIDILEVGLGKPTGKPTGKPTGKPTGKPTGKPTGKPT
GTSQGTGKGEKVPKGEQGLGPKLEGRVLAGSKGDKGKIGVIGKPGPTG
KGEKLOGFKGDTQPGIQQKGEKGEKGEKGEKGEKGEKGEKGEKGEKGEKGEKGEK
DRGKGEAGQRYKGEPPSQTINPAGKEQDRIKGEKNGITGKGTGKGTGKGTGKGT
SIGEPGKSGQKGTLDGPKGKQPAIGTPKGEKQDQGGPPGPGQDGIKEGK
DIDPGETDIDGPGIKGEFAGIKGTGPPGLOGPILGLKDDQGGPDAGKVPGR
SGDGFYLVKHSISIKVPSCAPAGMOTMBEGYSFLYAQNERAFQDLGQSGCLPKR
FSTWPLFCDIQNVASRNDYSFWSLTAEPKPEAPSSGADLENYTSRCLVCAPSH
VLAVHSELDKPCPDGMENLWTFSLMYNSAGAGSGQLLSGSGCLEDFRVNPYI
ECHRGTCWYGYPTLSFWLSTIGSNMFQVPKFEILLRNKARVSRCAVCKSVP"
157. .228
229. .276
/note="Region: non-collagenous domain"
277. .4641
/note="Region: contains collagenous domain"
526. .528
/note="N-glycosylation"
/evidence=not_experimental
4642. .5325
/note="Region: non-collagenous domain"
BASE COUNT 1953 a 942 c 1638 g 1318 t
ORIGIN

Query Match 26.2%; Score 40.6; DB 3; Length 5851;
Best Local Similarity 64.2%; Pred. No. 0.12;
Matches 61; Conservative 0; Mismatches 34; Indels 0; Gaps 0;
Qy 55 GGACCTCCAGCACACCCAGGCCGCCAGAGAGTGGGGCTCTCTGTCGACACAGTTTA 114
Db 2065 GGACCTCAAGGACTTCAAGGTCTCAAGGTGATCTGCTGCTCTCTGTTAAATCAGGTATA 2124
Qy 115 CCACAATATACAGGAGAAATAGTCAAAATGACAAA 149
Db 2125 CCAGCAATAACGAGAGAGAGAGTGAAGAGGAAA 2159

RESULT 3

HS438L4/c

LOCUS

DEFINITION

Human DNA sequence from clone 438L4 on chromosome 1p36.2-36.3

ACCESSION

297635

VERSION

HTG.

KEYWORDS

297635.10 GI:4835272

SOURCE

Homo sapiens.

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

I (bases 1 to 87100)

Direct Submission

Submitted (08-JUN-1999) Sanger Centre, Hinxton, Cambridgeshire,

CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk

requests: clonerequests@sanger.ac.uk

On May 14, 1999 this sequence version replaced gi:4678473.

During sequence assembly data is compared from overlapping clones.

Where differences are found these are annotated as variations

together with a note of the overlapping clone name. Note that the

variation annotation may not be found in the sequence submission

corresponding to the overlapping clone, as we submit sequences with

only a small overlap as described above.

The following abbreviations are used to associate primary accession

numbers given in the feature table with their source databases:

Emi.: EMBL; Sw.: SWISSPROT; Tri.: TREMBL

IMPORTANT: This sequence is not the entire insert of clone 438L4.

It may be shorter because we only sequence overlapping sections

once, or longer because we arrange for a small overlap between

neighbouring submissions.

The true left end of clone 505B13 is at 87001 in this sequence.

This sequence has been finished according to sequence map criteria

as follows. An attempt is made to resolve all sequencing problems,

such as compressions and repeats, but not necessarily within known

annotated human repeat sequence elements (e.g. Alu). Where the

sequence is ambiguous, there is an annotation using the 'unsure'

feature key.

This sequence was generated from part of bacterial clone contigs of

human chromosome 1, constructed by the Sanger Centre Chromosome 1

Mapping Group. Further information can be found at

<http://www.sanger.ac.uk/HGP/Chr1>

438L4 is from the library RPi3 constructed at the Roswell Park

Cancer Institute by the group of Pieter de Jong. For further

details see <http://bacpac.med.buffalo.edu/VECTOR:pcypac2>.

Location/Qualifiers

1..87100

/organism="Homo sapiens"

/db_xref="taxon:9606"

/chromosome="1"

/map="p36.2-36.3"

/clone="RP3-438L4"

/clone_lib="RPCI-3"

33..480

/note="match: GSS B78720 clone 382B5"

555..637

/note="L2 repeat: matches 2427..2508 of consensus"

746..773

/note="77 copies 4 mer caca 93% conserved"

746..771

/note="13 copies 2 mer ca 96% conserved"

1352..2056

/note="L1MFC repeat: matches 256..976 of consensus"

2071..2227

/note="L1MC repeat: matches 1557..1716 of consensus"

2228..2526

/note="AluX repeat: matches 2..300 of consensus"

2527..2709

/note="L1MC repeat: matches 1716..1890 of consensus"

2719..2835

/note="L1ME repeat: matches 6042..6157 of consensus"

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

/note="MIR repeat: matches 78..130 of consensus"
3286..3506
/note="AluSg/x repeat: matches 84..304 of consensus"
3517..3704
/note="LTR41 repeat: matches 597..772 of consensus"
4053..4193
/note="LTR41 repeat: matches 1..135 of consensus"
4911..5196
/note="AluSg repeat: matches 1..284 of consensus"
complement(<5103..>5396)
/note="match: STS G07800"
5193..5450
/note="match: STS L30457"
5197..5248
/note="13 copies 4 mer taga 98% conserved"
5197..5246
/note="25 copies 2 mer ta 76% conserved"
5426..5780
/note="L2 repeat: matches 2335..2709 of consensus"
6800..6807
/note="Single clone region"
6901..6954
/note="27 copies 2 mer ta 80% conserved"
6902..6957
/note="14 copies 4 mer atat 80% conserved"
7493..7746
/note="MER33 repeat: matches 5..258 of consensus"
7747..7889
/note="L1PA8 repeat: matches 6022..6163 of consensus"
7890..7953
/note="MER33 repeat: matches 258..322 of consensus"
9177..9674
/note="L1MB5 repeat: matches 5650..6176 of consensus"
9783..9866
/note="MIR repeat: matches 60..148 of consensus"
10330..10532
/note="MIR repeat: matches 13..224 of consensus"
11250..11290
/note="L2 repeat: matches 2665..2705 of consensus"
11947..12101
/note="MIR repeat: matches 18..183 of consensus"
12352..12693
/note="L2 repeat: matches 2376..2707 of consensus"
complement(<13712..>14212)
/note="match: GSS AQ139995"
15401..15769
/note="MLTII repeat: matches 57..410 of consensus"
15905..16078
/note="MER5A repeat: matches 1..183 of consensus"
16546..16698
/note="MIR repeat: matches 46..190 of consensus"
17702..18055
/note="L2 repeat: matches 2385..2749 of consensus"
18228..18337
/note="MIR repeat: matches 9..120 of consensus"
complement(1963..20348)
/note="match: EST AA594177 clone IMAGE:1085527"
20428..20785
/note="L1MB3 repeat: matches 5813..6185 of consensus"
20855..20914
/note="30 copies 2 mer tg 80% conserved"
21820..22105
/note="AluX repeat: matches 1..286 of consensus"
22218..22368
/note="HALI repeat: matches 1427..1589 of consensus"
24465..24800
/note="2 copies 168 mer 79% conserved"
24561..24680
/note="60 copies 2 mer ca 60% conserved"
24727..24778
/note="26 copies 2 mer ca 71% conserved"
24779..24826
/note="12 copies 4 mer acac 98% conserved"

TITLE	Genetic identification, sequence, and alternative splicing of the
JOURNAL	Caenorhabditis elegans alpha 2(IV) collagen gene
MEDLINE	J. Cell Biol. 123 (1), 255-264 (1993)
PUBMED	94012964
REFERENCE	7691828
AUTHORS	2 (bases 1 to 9541)
TITLE	Sibley, M. H.
JOURNAL	Direct Submission
FEATURES	Submitted (15-JUN-1993) Sibley M. H., Northwestern University
source	Medical School, Cell, Molecular and Structural Biology, 303 E.
5' UTR	Chicago Ave., Chicago, Illinois, USA, 60611
intron	Location/Qualifiers
exon	1..9541
mRNA	/organism="Caenorhabditis elegans"
CDS	/strain="N2 wild-type"
	/db_xref="taxon:6239"
	187..695
	217..676
	/number=1
	677..718
	/number=2
	join(<696..718,936..990,1188..1250,1625..1724,2047..2235,
	2736..2911,2963..3040,3161..3268,3299..3409,3807..3923,
	3970..4134,4184..4455,4503..5364,5417..6089,6367..7181,
	7227..7458,7512..8303,8386..8849,8908..>9541)
	/product="a2(IV) collagen"
	/note="alternatively spliced transcript"
	join(696..718,936..990,1188..1250,1625..1724,2047..2235,
	2736..2911,2963..3040,3161..3268,3807..3923,3970..4134,
	4184..4455,4503..5364,5417..6089,6367..7181,7227..7458,
	7512..8303,8386..8849,8908..9000)
	/note="alternatively spliced transcript"
	/codon_start=1
	/product="a2(IV) collagen"
	/protein_id="CAA80536.1"
	/db_xref="GI:953172"
	/translation="MKQRAALPVLRILAILALLAVSYVQSQATCRDCSNRGCFCVGEK
	GSMGAPQGPQGGTIGIRGFPFEGLAGKGLGAQGGPPGVGKIDRGAVGVGFPFG
	NDGNGRPGEPGPGAPGWDGNCNTDAGPIGRPRGPPGMPFGPPGMDGLKGEPAI
	GYACAGKEDGGMGMPGLPGPSRGDYGPEKGDRTGNAGRPPGGEAGSGNPG
	IGSLGPGDGLSGVPPGPPREFTGSGIVGRPNCKEKGKGGGGQRGYP
	GNGLSQDGLPGMKGKGLSGPAGPRGKGRGNAGPPGFKGDRGLDGLGIPGLPG
	KQKAGYVGRDRGKSGPPGPGGGTNDGAPGPGPLGRPRGNPPGTDGYPGAPG
	PAGYIGNTGPPGLPGYNEGLPGKDGIPGAPGVSPGSGIPLPGPKGEYP
	RGTPTQSIPLGLKGRGLDAGPRKGENGLPGVPRGPGDNLGLPCAPGQRCAGP
	NGYDGRGVNGLCAPGCTKDRGCTCSACAPGTGKEXGLPGYSQPGQDRGLPGMP
	GPVGDAGDGLPGAPRPGSPGPGQDGFPLGKQKGEPTQTLIRPDPGYPPLKGEN
	GFPGQDGLPGPSGPPGPPGAPGYPEKGDAGLGLSGKPGQDGLPLGNKGEAG
	YGPQPGVDFGAKDGLGLPGTPTGLQGMGEPAPENQVNPAPPGQGLPLGPTKG
	EGGYPRGPEVGQPFPLGCMKGDGLPGPPGLPHGVPDGRKGFVGLGPIGPG
	KGDVGNPLGLNKGKGPVGPQPSGPFGLKGDAGLPLGCTPTGLGQGRFPG
	APGLKGDGLPLGSLQPGYPEKGDAGLPGVPRGSGPFGQDGLPGVPMKGEDGL
	PLGLPVTLKGDLAGQSGAPGLGAPGYPMKNAGIPGVPGFKDGLPLGLPLGN
	KPGKPGVPMGPTGPMKNGSLPLGRDGLSGVPMKGDGRFNGLPGKGEAGPAA
	RDQKGDAGLPGQPLGKPGQPSGLPGVPFKETGLPGYGOPGPGKGLPIGPA
	GRQCAPSGQDGLPGPMKGSSEYPCQDGLPGRDGLPGVPGOKGDLGOSQPLSG
	APGLDGQPGVPIRGDKQGLPIPGDRGMDGYPGQKGENGYPCQGLPLGLGKGF
	AGTFFGLGSGYTPGQDGLPIGLKGDGSGFPQPGQEGPLGLSGEKMGPGMP
	GQPGQSTAGVPAGLQAGKQDGFPLGKQKESGLSPGLGAPGLKGESMPGFGA
	KDGLGANGIPGKRGDGLPGVPRDQGLPIGLKGEVAGLPGQPGFPIPLKGGEG
	GLPGFPAKGEAGFPFGVYAGEKGDGLPLGRDGLPGADGVPDPGPPGSPON
	LVEPEKGLPLCAPGLRGEKGMPLDGPNGDPPGLPCQGRNDGYPGAPGLSGEK
	GMGGLPFGDLDQGGPGAPPLGAPGAAGAPYRDGFVLVKHSQTTEVPCPGQTK
	LWDGYSLLYTEGNEKSHNDLGHAGSCLORFSTMPFLFCDFNNVCNYSRNEKSYWL
	TSEAIPMNVNERIEPYISRCACEAPTIAVHSQTIQIPNCPAGMSSLIWISYA
	MHTGAGAEQSGSPSGSCLEDFRATPFTFECNARGSCHYFANKFSFWLTIDNDE
	FKVPESQTLKSGNLTRVSRQVCVKSTDGRH"
	join(696..718,936..990,1188..1250,1625..1724,2047..2235,
	2736..2911,2963..3040,3299..3409,3807..3923,3970..4134,
	4184..4455,4503..5364,5417..6089,6367..7181,7227..7458,
	7512..8303,8386..8849,8908..9000)
	/note="alternatively spliced transcript"
CDS	

```

exon          /number=11
              3970. .4134
intron        /number=12
              4135. .4183
exon          /number=12
              4184. .4455
intron        /number=13
              4456. .4502
exon          /number=13
              4503. .5364
intron        /number=14
              5365. .5416
exon          /number=14
              5417. .6089
intron        /number=15
              6090. .6366
exon          /number=15
              6367. .7181
intron        /number=16
              7182. .7226
exon          /number=16
              7227. .7458
intron        /number=17
              7459. .7511
exon          /number=17
              7512. .8303
intron        /number=18
              8304. .8385
exon          /number=18
              8386. .8849
intron        /number=19
              8850. .8907
exon          /number=19
              8908. >9541
              /number=19
BASE COUNT   2776 a 2182 c 2252 g 2331 t
ORIGIN
Query Match      24.4%; Score 37.8; DB 3; Length 9541;
Best Local Similarity 55.8%; Pred. No. 0.93;
Matches 72; Conservative 0; Mismatches 57; Indels 0; Gaps 0;

QY 21 AATAAGCGGAGTCATTGAGCATGTAGAGTTGCTGGACCTCCAGCACACCCAGGCCCCC 80
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 5491 AATGAAGGGAGACTCTGGCTCCAGGACACCCAGGACTCCAGGACACCCAGGTGTTC 5550

QY 81 AGAAGAGTGGGGCCCTCCTGGTGCACAGGTTTACCACAAATATACAGGAGAAATAGTGA 140
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 5551 AGGAGACAAGGGATTTCGGAGGTGTCCAGGATTTACCGAGGAATTCAGGACCAAGGGAGA 5610

QY 141 AATGACAAA 149
    ||| |||
Db 5611 TGTTCGGAAA 5619

RESULT 8
U53342/c
LOCUS          U53342
DEFINITION     Caenorhabditis elegans cosmid F01G12, complete sequence.
ACCESSION      U53342
VERSION         U53342.1 GI:1255873
KEYWORDS       HTG.
SOURCE          Caenorhabditis elegans.
ORGANISM       Caenorhabditis elegans
               Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
               Rhabditoidae; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE      1 (bases 1 to 34671)
AUTHORS        Waterston,R.
TITLE          Genome sequence of the nematode C. elegans: a platform for
               investigating biology. The C. elegans Sequencing Consortium
               Science 282 (5396), 2012-2018 (1998)
JOURNAL        99069613
MEDLINE        9851916
PUBMED

2 (bases 1 to 34671)
REFERENCE      Wu,X. and Le,T.T.
AUTHORS        The sequence of C. elegans cosmid F01G12
JOURNAL        Unpublished (2001)
REFERENCE      3 (bases 1 to 34671)
AUTHORS        Waterston,R.
TITLE          Direct Submission
JOURNAL        Submitted (01-APR-1996) Robert Waterston
REFERENCE      4 (bases 1 to 34671)
AUTHORS        Waterston,R.
TITLE          Direct Submission
JOURNAL        Submitted (28-JUN-2001) Department of Genetics, Washington
               University, Genome Sequencing Center, 4444 Forest Park Avenue, St.
               Louis, MO 63110, USA
REFERENCE      5 (bases 1 to 34671)
AUTHORS        Waterston,R.
TITLE          Direct Submission
JOURNAL        Submitted (22-MAY-2002) Department of Genetics, Washington
               University, Genome Sequencing Center, 4444 Forest Park Avenue, St.
               Louis, MO 63110, USA
REFERENCE      6 (bases 1 to 34671)
AUTHORS        Waterston,R.
TITLE          Direct Submission
JOURNAL        Submitted (18-JUN-2002) Department of Genetics, Washington
               University, Genome Sequencing Center, 4444 Forest Park Avenue, St.
               Louis, MO 63110, USA
COMMENT        Submitted by:
               Genome Sequencing Center
               Department of Genetics, Washington University
               St. Louis , MO 63110, USA, and
               Sanger Centre, Hinxton Hall
               Cambridge CB10 1RQ, England
               email: rwenematode.wustl.edu and jesesanger.ac.uk

NOTICE: This sequence may not be the entire insert of this clone.
It may be shorter because we only sequence overlapping sections
once, or longer because we provide a small overlap between
neighboring submissions.

This sequence was finished as follows unless otherwise noted: all
regions were double stranded, sequenced with an alternate chemistry
or covered by high quality data (i.e., phred quality >= 30); an
attempt was made to resolve all sequencing problems, such as
compressions and repeats; all regions were covered by sequence from
more than one ml3 subclone.

For a graphical representation of this cosmid sequence and its
analysis see:
http://www.wormbase.org/db/seq/sequence?name=F01G12;class=Sequence

NEIGHBORING COSMID INFORMATION

The 5' cosmid is W09B12, 3100 bp overlap: the 3' cosmid is T24D11,
200 bp overlap. Actual start of this cosmid is at base position 1
of F01G12; actual end is at 34671 of F01G12.

NOTES:
Coding sequences below are the result of integration and manual
review of the following data : computer analysis using the program
Genefinder (P. Green and L. Hillier, personal communication), the
large scale EST projects of Yuji Kohara
(http://www.ddbj.nig.ac.jp/c-elegans/html/CE_INDEX.html) and The C.
elegans ORFeome cloning project (http://wordb.dfci.harvard.edu/),
similarity to other proteins from Blastx analyses
(http://blast.wustl.edu/), sequence conservation with C. briggsae
using Jim Kent's WABA alignment program (Genome Research
10:1115-1125, 2000), individual C. elegans GenBank submissions,
and personal communications with C. elegans researchers. tRNAs
are predicted using the program tRNAscan-SE (Lowe, T.M. and
Eddy, S.R., 1997, Nucl. Acids. Res., 25, 955-964).
               Location/Qualifiers
               1. .34671
FEATURES
source

```


Db	295	GCACCTGGTTTACC	282
RESULT 10			
MMCOLA4			
LOCUS	MMCOLA4	3612 bp	mRNA linear ROD 10-FEB-1999
DEFINITION	Mouse mRNA for basement membrane (type IV) collagen alpha chain.		
ACCESSION	X06777		
VERSION	X06777.1	GI:50501	
KEYWORDS	collagen; collagen alpha 1 type IV; collagen type IV.		
SOURCE	Mus musculus.		
ORGANISM	Mus musculus		
REFERENCE	1 (bases 46 to 3612)		
AUTHORS	Wood, L., Theriault, N. and Vogeli, G.		
TITLE	cDNA clones completing the nucleotide and derived amino acid sequence of the alpha 1 chain of basement membrane (type IV) collagen from mouse		
JOURNAL	FEBS Lett. 227 (1), 5-8 (1988)		
MEDLINE	88112221		
PUBMED	3338568		
REFERENCE	2 (bases 1 to 1000)		
AUTHORS	Wood, L.		
TITLE	Direct Submission		
JOURNAL	Submitted (02-SEP-1988)		
COMMENT	See M1402 for overlapping sequence (cDNA clone pcIV-1-225, same library).		
FEATURES			
source	1..3612		
	/organism="Mus musculus"		
	/db_xref="taxon:10090"		
	/clone="pcIV-1-PE12, pcIV-1-PE16"		
	/cloned_lib="pys cDNA"		
	75..76		
old_sequence	/note="uc was cc in [1]"		
	/citation=[1]		
old_sequence	131		
	/note="c was deleted [1]"		
	/citation=[1]		
CDS	151..23612		
	/note="collagen alpha chain precursor (AA -27 to 1127)"		
	/codon_start=1		
	/protein_id="CAA29946.1"		
	/db_xref="GI:50502"		
	/db_xref="MGD:MG1:88454"		
	/db_xref="SWISS-PROT:P02463"		
	/translation="MGPRLSVWLILLFAALLLHEERSRAPAKGDCGGSGCGKDCDHGV		
	KQXGERGLPCLQVIGFGMCGPEGPHGPGGQKGDAGEPLGCTKGTGRPPCAAGY		
	CNPLGLPIPGDGPDPGPGIPGCGNTKGERGLPGPLGFCFSNRPGLPGMKGDPG		
	ETLHVPTLLKGERGFPPIGMPGLPGLQGPVGPPTGPPGPPGPPGPPGK		
	QMSFSGQPKGDKGEQVSGVPQQAQYKEKGDAPTGEKQKGEPPGPPGPPG		
	KGERKQGRGKPKGDKGERGSPGIFGDSGYPGLGRQPSGERGEAGLPGPPTG		
	IGTWPLGKGRGYPGAPLGRGEPGKGFPTGLPGPPTPGAGAGPFGERGE		
	KGDGFCVSPFPSPGRDAGPDPGPPGPPGPPGPPGPPGPPGPPGPPG		
	PGLGVEVCQKQKGESCLADTEGLRGPPGLQGPPEIGFPHPGAKGDRGLPGRGL		
	EGLPQSGSLIGQCAKGEPEIFFDMRLKDKDGPDPGPPGPPGPPGPPGPPG		
	PGLFPGKSPGSLGLKGERPPGPGVGGGIPGPPGPPGPPGPPGPPGPPGPPG		
	GPSPGLPFGKGEAGVPLPDPGAAAGLPGSPGFPQGDGRGFPCTGPPRPIPGK		
	AVGQPIGFPLGPKGVVDGLPGLGIRGPIPRGPNGLPFPNGPFGQKGEPIGLPL		
	LKQPLGLPIPTGEXKSGTGGPVGPEQGLTGPPGLOGIRGDPGPPGVGPPGPPG		
	FGIIPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG		
	SGLPGLGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG		
	GPGRDPPGFKGDGDVGLPFGMPGSMHVDGMSMKQKQDKQKQIGTGDGKSGRGP		
	TPGVGDKGHAGHPQPKDGLSGTSPGSLPGPKGVPKSGVMGMLGPPGSGKGVPI		
	PGSQGVPSGCEKCAKGEKSGSLPGIIPGRPDKDGCLAGFPGSGKGEKGSAG		
	TPGMFSGPRGSPGNIGHGSPGLPGEKDKGLPGLDGLGVGKGEAGLPGTGPPTG		
	AGQKGEPPG"		
	151..231		
	232..1882		
sig_peptide			
mat_peptide			

[illegible]

Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N., Nguyen, N., Nickerson, E., Nwokenkwo, S., Oguh, M., Okwuonu, G., Oragunye, N., Oviado, R., Pace, A., Payton, B., Peery, J., Perez, L., Peters, L., Pickens, R., Primus, E., Pu, L.L., Quiles, M., Ren, Y., Rives, M., Rojas, A., Rojubokan, I., Roife, M., Ruiz, S., Savery, G., Scherer, S., Scott, G., Shen, H., Shoostari, N., Sisson, I., Suttongren, E., Sonaik, T., Sparks, A., Stanley, H., Stone, H., Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H., Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S., Usmani, K., Vasquez, L., Vera, V., Villalon, D., Vinson, R., Wang, O., Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S., Williams, G., Williamson, A., Wleczyk, R., Wooden, S., Worley, K., Wu, C., Wu, Y., Wu, F., Zhou, J., Zorrilla, S., Nelson, D., Weinstein, G. and Gibbs, R.

Direct Submission
Unpublished
2 (bases 1 to 95492)
Worley, K.C.

Direct Submission
Submitted (17-SEP-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 95492)
Worley, K.C.

Direct Submission
Submitted (11-JUL-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
On Jul 10, 2002 this sequence version replaced gl:17943819.

----- Genome Center -----
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu

----- Project Information -----
Center project name: GEM
Center clone name: CH230-11E5
----- Summary Statistics -----
Sequencing vector: plasmid
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 36046 bases at least Q40
Consensus quality: 42440 bases at least Q30
Consensus quality: 46490 bases at least Q20

* NOTE: Estimated insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 46 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

* 1 1132: contig of 1132 bp in length
* 1133 1232: gap of unknown length
* 1233 2500: contig of 1268 bp in length
* 2501 2600: gap of unknown length
* 2601 4032: contig of 1432 bp in length
* 4033 4132: gap of unknown length
* 4133 5210: contig of 1078 bp in length
* 5211 5310: gap of unknown length
* 5311 6685: contig of 1375 bp in length
* 6686 6785: gap of unknown length
* 6786 8171: contig of 1386 bp in length
* 8172 8271: gap of unknown length
* 8272 9447: contig of 1176 bp in length
* 9448 9547: gap of unknown length
* 9548 11015: contig of 1468 bp in length
* 11016 11115: gap of unknown length
* 11116 12931: contig of 1816 bp in length
* 12932 13031: gap of unknown length

* 13032 14745: contig of 1714 bp in length
* 14746 14845: gap of unknown length
* 14846 16501: contig of 1656 bp in length
* 16502 16601: gap of unknown length
* 16602 17952: contig of 1351 bp in length
* 17953 18052: gap of unknown length
* 18053 19788: contig of 1736 bp in length
* 19789 19888: gap of unknown length
* 19889 21580: contig of 1692 bp in length
* 21581 21680: gap of unknown length
* 21681 22901: contig of 1221 bp in length
* 22902 23001: gap of unknown length
* 23002 24262: contig of 1261 bp in length
* 24263 24362: gap of unknown length
* 24363 25367: contig of 1005 bp in length
* 25368 25467: gap of unknown length
* 25468 25915: contig of 1448 bp in length
* 25916 27015: gap of unknown length
* 27016 28543: contig of 1528 bp in length
* 28544 28643: gap of unknown length
* 28644 30278: contig of 1635 bp in length
* 30279 30378: gap of unknown length
* 30379 31382: contig of 1004 bp in length
* 31383 31482: gap of unknown length
* 31483 33124: contig of 1642 bp in length
* 33125 33224: gap of unknown length
* 33225 34854: contig of 1630 bp in length
* 34855 34954: gap of unknown length
* 34955 36580: contig of 1626 bp in length
* 36581 36680: gap of unknown length
* 36681 37765: contig of 1085 bp in length
* 37766 37865: gap of unknown length
* 37866 39561: contig of 1696 bp in length
* 39562 39661: gap of unknown length
* 39662 41262: contig of 1601 bp in length
* 41263 41362: gap of unknown length
* 41363 43753: contig of 2391 bp in length
* 43754 43853: gap of unknown length
* 43854 47134: contig of 3281 bp in length
* 47135 47234: gap of unknown length
* 47235 48919: contig of 1685 bp in length
* 48920 49020: gap of unknown length
* 49020 51487: contig of 2468 bp in length
* 51488 51587: gap of unknown length
* 51588 53919: contig of 2332 bp in length
* 53920 54019: gap of unknown length
* 54020 56113: contig of 2094 bp in length
* 56114 56213: gap of unknown length
* 56214 58669: contig of 2456 bp in length
* 58670 58769: gap of unknown length
* 58770 61080: contig of 2311 bp in length
* 61081 61180: gap of unknown length
* 61181 63791: contig of 2611 bp in length
* 63792 63891: gap of unknown length
* 63892 66216: contig of 2325 bp in length
* 66217 66316: gap of unknown length
* 66317 68991: contig of 2675 bp in length
* 68992 69091: gap of unknown length
* 69092 71765: contig of 2674 bp in length
* 71766 71865: gap of unknown length
* 71866 73926: contig of 2061 bp in length
* 73927 74026: gap of unknown length
* 74027 76639: contig of 2613 bp in length
* 76640 76739: gap of unknown length
* 76740 80127: contig of 3388 bp in length
* 80128 80227: gap of unknown length
* 80228 83868: contig of 3641 bp in length
* 83869 83968: gap of unknown length
* 83969 88422: contig of 4454 bp in length
* 88423 88522: gap of unknown length
* 88523 92040: contig of 3518 bp in length
* 92041 92140: gap of unknown length
* 92141 95492: contig of 3352 bp in length.

FEATURES
source Location/Qualifiers
1. .95492
/organism="Rattus norvegicus"
/db_xref="taxon:10116"
/clone="CH230-11E5"
BASE COUNT 28940 a 16456 c 17230 g 28330 t 4536 others
ORIGIN
Query Match 23.6%; Score 36.6; DB 2; Length 95492;
Best Local Similarity 58.9%; Pred. No. 2.7; Indels 0; Gaps 0;
Matches 63; Conservative 0; Mismatches 44;
Qy 8 TCATACCTGCTTAAATAACGGCAGTCATTGACGATGTAGAACTTGGTGACCTCCAGCAC 67
||||| || | | | | ||||| ||||| | ||| ||
Db 22044 TCATCCCATACCTCCTCCCAAGGTCAAACAGGATGCTCTATGCTCTCTATGCTCTCCAA 21985
Qy 68 ACCCCAGGCCCCAGAAAGTGGGGCTCTCTGGTGACCCAGGTTTA 114
||||| ||||| || | | ||||| || | | ||||| |||
Db 21984 ACCCCAGGCTCCCTATCACCTGGGGCTCAAGTCTCTCAGGGTTA 21938
Search completed: February 20, 2003, 04:35:08
Job time : 694.184 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 19, 2003, 23:48:42 ; Search time 3481.23 Seconds
(without alignments)
6406.126 Million cell updates/sec

Title: US-09-997-610-1-copy_2_1378

Perfect score: 1377

Sequence: 1 atagtgtcatactgtctt.....gcctgtacctccattgtatg 1377

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*

- 1: em_estba:**
- 2: em_esthum:**
- 3: em_estin:**
- 4: em_estmu:**
- 5: em_estov:**
- 6: em_estpl:**
- 7: em_estro:**
- 8: em_htc:**
- 9: gb_est1:**
- 10: gb_est2:**
- 11: gb_htc:**
- 12: gb_est3:**
- 13: gb_est4:**
- 14: gb_est5:**
- 15: em_estfun:**
- 16: em_estom:**
- 17: gb_gss:**
- 18: em_gss_hum:**
- 19: em_gss_inv:**
- 20: em_gss_pln:**
- 21: em_gss_vrt:**
- 22: em_gss_fun:**
- 23: em_gss_mam:**
- 24: em_gss_mus:**
- 25: em_gss_Other:**
- 26: em_gss_pro:**
- 27: em_gss_rod:**

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	675.4	49.0	2615	10	BE420422
2	615	44.7	2009	11	BC004496
3	574.4	41.7	1514	10	BE512633
4	480.6	34.9	1036	13	BM471183
5	479.2	34.8	870	13	BI488505
6	477.8	34.7	898	13	BM457166

7	469.8	34.1	937	14	BQ893011
8	462.6	33.6	1050	13	AGENCOURT
9	451.4	32.8	941	14	BM472108
10	435.6	31.6	876	14	BQ723415
11	430.4	31.3	717	17	AGENCOURT
12	421.4	30.6	1076	13	BQ423563
13	419.8	30.5	676	17	AGENCOURT
14	416.4	30.2	1063	14	AG086951
15	414.8	30.1	1080	14	BM554723
16	411.4	29.9	654	17	AGENCOURT
17	411	29.8	721	12	BQ422427
18	407.8	29.6	955	12	AGENCOURT
19	405.4	29.4	760	12	AG112666
20	404.6	29.4	736	13	BE728616
21	404.2	29.4	691	17	BG681769
22	403.8	29.3	900	13	BG281182
23	403.2	29.2	998	14	B1754555
24	402.2	29.2	1182	13	AG062447
25	401.2	29.1	701	17	AG062447
26	401.2	29.1	704	17	BM449699
27	399.2	29.0	729	13	BM903785
28	398.4	28.9	720	17	BM451317
29	396.2	28.8	697	17	AQ588873
30	394.4	28.6	699	9	AG588828
31	394.2	28.6	833	9	BM045873
32	393.8	28.6	677	17	AG116938
33	393.2	28.6	877	13	AG018926
34	392.4	28.5	659	17	AG097258
35	390.8	28.4	639	17	BI916691
36	389.4	28.3	676	17	AG092858
37	387.8	28.2	623	17	AG115932
38	387.8	28.2	704	17	AG125093
39	386.6	28.1	771	17	AQ037711
40	386.4	28.1	885	12	AQ020199
41	383.4	27.8	671	17	AG001229
42	381	27.7	742	17	AG001229
43	380.6	27.6	660	17	BG761979
44	380.2	27.6	686	17	AG045131
45	380	27.6	937	14	AG045131

ALIGNMENTS

RESULT 1
BE420422 LOCUS BE420422 2615 bp mRNA linear EST 11-OCT-2001
DEFINITION 32-393 human bone marrow cDNA library Homo sapiens cDNA, mRNA
sequence.
ACCESSION BE420422
VERSION BE420422.1 GI:16041640
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniala; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 2615)
AUTHORS Lu,X., Cul,L. and Li,Y.
TITLE DDRT-PCR from B cell
JOURNAL Unpublished (2000)
COMMENT Contact: xingwu Lu, llangxian Cul,yonghal Li
Department of Biochemistry
Institute of Basic Medical Science, Peking Union Medical College
Dongdan Sanriao 5, Beijing, P.R.C, 100005
Tel: 86-010-65296951
Email: luxingwu@263.net
full-length and coding sequence.

FEATURES
source

Location/Qualifiers
1..2615
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="human bone marrow cDNA library"
/tissue_type="bone marrow"

	BASE COUNT	ORIGIN
1	0	0
2	0	0
3	0	0
4	0	0
5	0	0
6	0	0
7	0	0
8	0	0
9	0	0
10	0	0
11	0	0
12	0	0
13	0	0
14	0	0
15	0	0
16	0	0
17	0	0
18	0	0
19	0	0
20	0	0
21	0	0
22	0	0
23	0	0
24	0	0
25	0	0
26	0	0
27	0	0
28	0	0
29	0	0
30	0	0
31	0	0
32	0	0
33	0	0
34	0	0
35	0	0
36	0	0
37	0	0
38	0	0
39	0	0
40	0	0
41	0	0
42	0	0
43	0	0
44	0	0
45	0	0
46	0	0
47	0	0
48	0	0
49	0	0
50	0	0
51	0	0
52	0	0
53	0	0
54	0	0
55	0	0
56	0	0
57	0	0
58	0	0
59	0	0
60	0	0
61	0	0
62	0	0
63	0	0
64	0	0
65	0	0
66	0	0
67	0	0
68	0	0
69	0	0
70	0	0
71	0	0
72	0	0
73	0	0
74	0	0
75	0	0
76	0	0
77	0	0
78	0	0
79	0	0
80	0	0
81	0	0
82	0	0
83	0	0
84	0	0
85	0	0
86	0	0
87	0	0
88	0	0
89	0	0
90	0	0
91	0	0
92	0	0
93	0	0
94	0	0
95	0	0
96	0	0
97	0	0
98	0	0
99	0	0
100	0	0

Qy	351	GAATAATTGGCTAATGAGGAACAATAATTTGGCTAATAAGGAAGAAATTTCTAAGCAGCA	410
Db	493	GAACCTTTAGCTTTGAGAGAGATCATTTTGGGTATCTGGCAGAAGAAATTTCTAAGCAGCA	552
Qy	411	AAGCATTCAAGAGGTACATTTGGTGCTGTAAAGGCATTCAGTTTCATAAAGGGAGGCAGA	470
Db	553	AAGCATTCAAGAGGTAACTTTGGGTGCTGTAAAGGCATTTAGTTTATTAAGGAAGCAGG	612
Qy	471	GCATAAGAGTTCAGAAATTTGCACCTGCACAAATGTGATAAAAGAAAGAAACCCATTTC	530
Db	613	GCATGAAGTTTGGAAATTTGCACCTGCACAAATGCATAGAAAGAAATCCCATATTC	672
Qy	531	TGAGGGAAATTCAGCTGGCTGCAGAAATTTGCATATGTAAATGAGGAGCTGAATGTAA	590
Db	673	TGAGGAGAAAAATCAAGCTGGCTGCAGAAATTTGCATATGTAAACAAAGAGTTAAATGTAA	732
Qy	591	TCCTCAAGACAATGGGGAATAATCTCTGGACATGTCAGAGGTCTTCACAGCAGTCCAT	650
Db	733	TCCTCAAGACAATGGGGAATAATCTCTCAGGSCATGCCAGAGGTCTTCACAGCAGCCCT	792
Qy	651	CAAAATCACTGGCTGGAGGCTTAGGAG---AAAATGGTTTTGTGGACAGGCCACAGGTC	707
Db	793	CCCATACAGGCTGCAGGCTTAGGAGGAAAAGTGATCTTGGGCGAGGCCACAGGCT	852
Qy	708	CTGTGCTGTGTACGCTTAGAGACTTGTGGCTGTGTCCAGTTAATTCAGC-----	761
Db	853	CCCATGCTGTGCAGCTTGGGACTTGGTGCCTTGTGCCAGCTGCTCCAGCCATGGC	912
Qy	762	-----TGTTGCTTCAGAGGTCGAAGCCCTGGATGCCAGGCAAGATTTGC	914
Db	913	TGAAAGGGGCCAAATAGAGCTCGGATTTGGCTTCAGAGGTCGAAGCTGAAGCCCTG	972
Qy	795	GCAGCTTCCAAGTGTGTTGACCTGTGGTGTCCAAGAGTCAAGAATTGAGGTTGGGA	854
Db	973	GCAGCTTCCAATGTGGTGTGAGCTTGCCAGTGCACAGAAATCAAAAATTTGGGTTTGGGA	1032
Qy	855	ACCTCCAATCAGATTTCAAGAATATATGGAACCCCTGGATGCCAGGCAAGATTTGC	914
Db	1033	ACCTCTGCCTAGATTTCAGAAATGATGGAATGCCCTGGATGCCCTGGCAGAGTCTGC	1092
Qy	915	TGTAGGGGTGGGTCTCATGAGAAACCTCTGCAAGGTTAGTACAAAAGGAAATGTTGG	974
Db	1093	TGTAGGGAGGGGCCCTCATGGAGAATCTCTGCTAGGCGAGTACAGAAGGCAAAATGTTGG	1152
Qy	975	GTGGGAGCCCCACACAGAGTCCCCAGTGGGGCTCCATCTAGTAGAGCTGTGAGAAGAAG	1034
Db	1153	ATCGGAGCCCCACAGCAAGATCCCTACTGGGGCACCCGCTAGTGGAGGCTGTGAGAAGAG	1212
Qy	1035	TCCACCATCTCCAGACTCCAGAAGGTTAGATCCACTGCACAGCTTGACAGATGTCCTGA	1094
Db	1213	GCCACCATCTCCAGACCCCAAGATGATATCCACTGACAGCTTGTCCATGCACCTGG	1272
Qy	1095	AAAATCCACAGACTCAGTGGCAGCCTGTGAAAGCAGCAGGGATGAGTCTGTACCCCTA	1154
Db	1273	AAAACCGCAGA---CAACACCGCCCATGAACACAGCTAGGCAGGAGGCTGTACCCCTG	1328
Qy	1155	CAAAACCGTAGTGGCAGAGCTGACCAAGACCGTGGGAATCTACCTCTTGCAATGTCATGA	1214
Db	1329	CAAAAGCCACAGGGGGGAGCTGCCAAGATCATGGGAACCCACCTCTGTGCATCAGCATGA	1388

REMARK NIH-MGC Project URL: <http://mqc.nci.nih.gov>

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/IMG at: <http://image.llnl.gov>
 Series: IRAL Plate: 14 Row: d Column: 7
 This clone was selected for full length sequencing because it passed the following selection criteria: Hexamer frequency ORF analysis

analysis
This clone has the following problem: frame shifted.

```

source
1..2009
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:381313"
/tissue_type="Skin, melanotic melanoma."
/clone_lib="NIH_MGC_20"
/lab_host="DH10B-R"
/note="Vector: pOTB7"
556 a 429 c 513 g 511 t
BASE COUNT
ORIGIN

```

Query Match	44.7%	Score 615;	DB 11;
Best Local Similarity	78.2%	Pred. No. 1.9e-170;	
Matches 781;	Conservative	0;	Mismatches 205;
			Indels 13;
			Gaps 3;
			Length 209;

391 GAAGAAATTTCTAAGCAGCAAGCATTCAAGAGTGACTTGGTGCTGTTAAAGGCATTTC 450
|||||
490 GAAGAAATTTCTATGCAAGCAAGCATTCAAGAGTTGACTTGGTGCTGTTAAAGGCATTTC 549
|||||
451 AGTTTCATAGGGAGGCAGACATAAGAGTTTCAGAAATTTGCACCCCTGCACATGTCGATA 510


```
Db 550 AGATTATAAGGGAGCAGACATAAAAGTTTCAGAAAAATTTGACGCTTGACAATGTGATA 609
Qy 511 AAAAAGAAAAACCCATTTCCTGAGGGGAAATTCAGCTGGCTGCAGAAATTTGCAATATGT 570
Db 610 GAGAGAAAAAGCCATTTTGTGAGGAGAAATTCAGCCAGCTGCATAATTTTGCATAAGT 669
Qy 571 AATGAGGAGCTGAATGTTAACTCCTCAAGACAATGGGAAAAATATCTCTCGACATGTGAC 630
Db 670 AACAGAAGCGCATGTTAACTCCCAAGACAACGAGAGACATCTCCAGGACATATCAG 729
Qy 631 AGCTTTCACAGCAGCTCATCAATCACTAGCTGGCTGGAGCCCTAGAGAAAAATGTTTGT 690
Db 730 AGGCTTTCACAGCAGCCCTCCCTGCACAGGCCCAAGAGCCCTAGAGAAAAATGTTTCT 789
Qy 691 GGGAC-----AGCCAGGCTCCTGTGTGTGTGTCGACGCTAGACAGCTTGTGCTCC 740
Db 790 GGACCTGTGCAGCCTAGTACTTGGCACCTCTGTCCAGACTACTCCAGTAGTGCTGAA 849
Qy 741 CTGTGTCCAGTTAA--TTACGCTGTGCTTTCAGAGGCTGCAAGCCCAAGCCTTTGGCAG 798
Db 850 AGGGCCAAAGTACAGCTTACGCCATTACGCCCTTTCAGAGGCTGCAAGCCCAAGCCTTTGGCAG 909
Qy 799 CTTCCAAGTGTGTGAGCCTGTGGTGCAGAAAGTCAAGAAATGAGGTTTGGGAACCT 858
Db 910 CTTCCGATGCTGTGTGAGCCTGCAAAATGCACAGAGCTCAAGAAATTTGGATTGGGAACCT 969
Qy 859 CCAATCAGATTTCAGAACATATATGGAACCCCTGGATGCCAGCAGCAGAACTTGTCTCTA 918
Db 970 CTGCTAGATTTCAGAGAGATGTACGGAAGTGTCCGGATGCCCAAGTTTGTCTCA 1029
Qy 919 GGGGTGGGCTCCTCATGAGAACCTCTCAAGGCTAGTACAAAAGGGAATTTGGGTGG 978
Db 1030 GGGCAGGGCTTTCATGAGAACCTCTGCTAGGCGAGTGTGGAAGGGAACATGGGCT 1089
Qy 979 GAGCCCCACACAGAGTCCCCAGTGGGCTCCATCTAGTAGAGCTGTGAGAGAAAGTCCA 1038
Db 1090 GAGCCCCACACAGAAATCCCTACTGAGGCACACCTAGTGAGGCTGTGAGAAAAAGAGCA 1149
Qy 1039 CCATCCTCAGACTCCAGAGGCTAGATCCACTGACAGCTTTCAGCAGTGTGCTGCAAAA 1098
Db 1150 CCATCCTCCACCCCAAGATATAGATCCACACAGCTTGCACCTGCACTGGAAAA 1209
Qy 1099 TCCACAGACTCAGTGCCAGCCTGTGAAAGCAGCAGGAGTGGAGTCTGTACCCCTACAAA 1158
Db 1210 GCCGACAGATTCAACACACGCCCTGAAAGCAGCCAAAGGGGAATATATCTTGCAA 1269
Qy 1159 ACCGTAGTGAGAGCTGACCAAGACCGTGGGAATCTACCTCTTGTGCTGATGATGACCTG 1218
Db 1270 GCCACAGGGAGGAGCTTCCCAAGACTATGGGAACCTACCTCTTACATCAGCAGACCTT 1329
Qy 1219 GACGTGAGACATGGAGTCAAAAGAGATCATTTTGGAGCTTTAAGATTGACTGCCCCACT 1278
Db 1330 TGTGTGAGACCTGGAGTCAAAAGAGATCATTTTGGAGCTTTTAAATCTGACTGCCCCCT 1389
Qy 1279 GGATTTGGAGCTTATATGGGGCC--CGTACCCCTTTGTTTGGCCAAATTTTTCATTTGG 1337
Db 1390 GGATTTGAGACTTGATGGGGCCCTACACCCCTTTGTTTGGCCGATTTCTCCAATTTGG 1449
Qy 1338 AACTGCCGATTTTACCCCAATCCCTGTACCTCCATTTGAT 1376
Db 1450 AAAGCTGTATTTACCCCAATACCTGTACCCCATTTGAT 1488
```

```
RESULT 3
BE512633
LOCUS 32-1514 human B lymphocyte cDNA library Homo sapiens cDNA, mRNA
DEFINITION BE512633
ACCESSION BE512633
VERSION BE512633.1 GI:16041645
KEYWORDS EST.
SOURCE human.
```

```
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1514)
AUTHORS Lu,X., Cui,L. and Li,Y.
JOURNAL DDRT-PCR from B cell
COMMENT Unpublished (2000)
Contact: xingwu lu,liangxian cui,yonghai li
Institute of Basic Medical Science, Peking Union Medical College
Dongdan Sanliiao 5, Beijing, P.R.C. 100005
Tel: 86-010-65296951
Email: luxingwu@263.net
full cDNA sequence.
FEATURES
Location/Qualifiers
source 1..1514
/organism="Homo sapiens"
/db_xref="taxon:9606"
/tissue_type="bone marrow"
/note="Organ: tongue; Vector: pAMP10; mRNA made from
tongue epithelium, cDNA made by oligo-dT priming.
Non-directionally cloned into UDG sites. Size-selected on
agarose gel, average insert size 500 bp. Primary library.
cDNA Library Preparation: David B. Krizman, Ph.D.
REFERENCE: Krizman et al. (1996) Cancer Research
56:5380-5383."
BASE COUNT 380 a 341 c 398 g 395 t
ORIGIN
Query Match 41.7%; Score 574.4; DB 10; Length 1514;
Best Local Similarity 81.0%; Pred. No. 1.8e-158;
Matches 734; Conservative 0; Mismatches 131; Indels 41; Gaps 4;
Qy 508 ATAAAAAGAAAAACCCATTTCCTGAGGGGAAATTCAGCTGGCTGCAGAAATTTGCATA 567
Db 1 ATAGAAAAAGAAATCCCATATTCCTGAGGAGAAATTCAGCTGGCTGCAGAAATTTGCATA 60
Qy 568 TGTATGAGGAGCTGAATGTTAACTCCTCAAGACAATGGGAAATATCTCCTGGACATGT 627
Db 61 TGTAAACAGGAGTTAAATGTTAAATCCCAAGACAATGGGAAATGTTCTCCAGGCGCATGC 120
Qy 628 CAGAGGCTTTCACAGCAGCTCCATCAATCACTGGCTTGAGGCGCTTAGGAGGAGGAGG 684
Db 121 CAGAGGCTTTCACAGCAGCCCTCCCATCACAGGCTTCAGGCGCTTAGGAGGAGGAGG 180
Qy 695 TTTTGTGGCAGGCGCCAGGCTCCCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 744
Db 181 TCTCTTGGGCCAGGCCAGGCTCCCATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 240
Qy 745 GTCCCAGTTAATTCAGC-----TGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 771
Db 241 GTCCCAGCTGTCCAGCCATGGCTGAAAGGGGCCAACATAGAGCTCGGATTTGTGCTTCA 300
Qy 772 GAGGTGCAAGCCCCCAAGCCTTGGCAGCTTCCAACTGGTGTGTGTGTGTGTGTGTGTGTGT 831
Db 301 GAGGTGCAAGCTGTAGCCTTGGCAGCTTCCATGTGTGTGTGTGTGTGTGTGTGTGTGTGT 360
Qy 832 AAGTCAAGAAATGAGGTTTGGGAACCTCCAATCAGATTTCAAGAAATATATGGAACCC 891
Db 361 AAGTCAAAATTTGGGTTTGGGAACCTCTGCTAGATTTCAAGAAATGATATGGAATGCC 420
Qy 892 TGGATGCCCAGGCAGAACTTTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 951
Db 421 TGGATGCCCTGGCAGAACTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 480
Qy 952 GTAGTACAAAAGGGAAATTTGGGTGGGAGGCCCCACACAGAGTCCCGAGTGGGCTTCCA 1011
Db 481 GCAGTACAGAAGGAAATGTGGGATCGGAGGCCCGCCAGCAGACAGAAATCCCTACTGGG 540
Qy 1012 TCTAGTAGAGCTGTGAGAAGAGTCCACCATCTCTCCAGACTCCAGAGGGTAGATCCACT 1071
Db 541 CCTAGTGGAGCTGTGAGAAGAGGGGCCACCATCTCTCCAGACCCAGAAATGATAGATCCACT 600
```


AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>;
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabbs-r@mail.nih.gov
Tissue Procurement: Life Technologies, Inc.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: LLAM1479 row: k column: 04
High quality sequence stop: 849.
Location/Qualifiers
1. .870
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5191803"
/clone_lib="NIH_MGC_114"
/lab_host="DH10B"
/note="Organ: brain; Vector: pCMV-SPORT6; Site 1: NotI; Site 2: EcoRV (destroyed); RNA source anonymous pool of 6 male brains, age range 23-27 yo. Library is oligo-dT primed and directionally cloned (EcoRV site is destroyed upon cloning). Average insert size 1.5 kb, insert size range 1-3 kb. Library is normalized and enriched for full-length clones and was constructed by C. Gruber (Invitrogen). Research Genetics tracking code 019. Note: this is a NIH_MGC Library."
BASE COUNT 222 a 211 c 256 g 181 t
ORIGIN

Query Match 34.8%; Score 479.2; DB 13; Length 870;
Best Local Similarity 76.1%; Pred. No. 2.3e-130;
Matches 645; Conservative 0; Mismatches 168; Indels 35; Gaps 3;

Qy 492 GCACCCTGACATGTGATAAAGAAAGAAACCCCATTTCTGAGGGGAAATTCAGCTGGC 551
Db 1 GCAGCCTGATCATGAAGTAGAAGAAAGAAACCCCATTTTTCAGGATAAATTCAGCCAGC 60

Qy 552 TGCAGAAATTCGATATGTAATGAGGAGCTGAATGTTAATCCTCAAGACAAATGGGGAAA 611
Db 61 TGCAGAAATTCGATATGTAACAGAGGAGCTGAATGTTAATCCTCAAGACCAATGGGGAAA 120

Qy 612 TATCTCTGGACATGTCAGAGGCTCTCACAGCACTCCATCAATCACTGGCTTGAGGC - 670
Db 121 TGTCTCGGGCCATGTCAGAGACCTTCATGGCAGCCCTCCCATCAGAGGCTGGAGGCCA 180

Qy 671 CTAGGAGAAATGGTTTGTGGGACAGCCAGGCTCCCTGCTGTGTGAGCCTAGAG 730
Db 181 GGAGAAAGAGGTTTCGTGGCCAGCCAGCATCCCCGTGCTTGTGAGCCTAGG 240

Qy 731 ACTTGGTCCCTGTGTCACGATTAAATCA----- 759
Db 241 ACTTGGTCCCTGTGTCACGCTCCAGCTGTGACTGAAAGGGGCCAACGTAGAGTC 300

Qy 760 --GCTGTGGCTTCAGAGGTCGAAGCCCAAGCCCTGCGACCTTCCAGTGGTTGAGC 817
Db 301 GGGCTGTGGCTTCAGAGGTTGGAAGCTCTAAGCTTTTGGCAGCTTCCACGTGGTTGAGC 360

Qy 818 CTGTGGTGGCAAGAAGTCAAGAAATAGAGTTTGGGAACCTCCCAATCAGATTTCAGAGA 877
Db 361 CTACAGGTGCACAGAGTCAAGAACTGAGGTTTGGGAACCTCTCCCTAGATTTCAGAGA 420

Qy 878 TATATGGAACCCCTGGATGCCAGGAGAAAGTTTGTGTAGGGTGGGCTCCTCATGGA 937
Db 421 TGTGTGAAATGCCTGGATGCCAGGAGAAAGTTTGTGTGACAGAGTGGGGCCCTTATGGA 480

Qy 938 GAACCTCTGCAAGGTAGTCAAAAGGAAATGTTGGTGGAGCCGCCCCACACAGAGTCC 997
Db 481 GAACCTCTGCTGGAGCAGTGCAGAAAGGAAATGTGGATCGGAGCCGCCCCACAGAGTCC 540

Qy 998 CCAGTGGGGCTCCATCTAGTAGAGCTGTGAGAAGAGTCCACCATCTCCAGACTCCAGA 1057
Db 541 ATACTGGGGCACTGCCTAGTGGAACTGTGAGAAGACACACCCGCTCTCCAGACCCAGA 600

Qy 1058 AGGTGGATCCACGTACAGCTTGCAGCATGTGCCTGAAAAATCCACACACACTCAGTGCC 1117
Db 601 ATGGTAGATCCACCAACACAGCTTGCACCATGTGCCTGGAAAAGATGCAGGTACTCAATGCC 660

Qy 1118 AGCCTGTGAAAGCAGCAGGATGGAGTGTGATACCTCAACCAACCGTAGTGGCAGAGCTGA 1177
Db 661 AGCCCATGAAGCAGCCAGGAGGAGGCTGTACCTCTGCAAGCCACACAGGGAGGAGTGT 720

Qy 1178 CCAAGACCGTGGGAATCTACCTCTTGCATTTGCATGTGATGACCTGGACGTGAGAGATGAGTCA 1237
Db 721 CCAAGACCATGGAAACCCACTCTTGCATCAGCGTGACCTGGATGCGAGACCTGGAGTCA 780

Qy 1238 AAAGAGATCAATTTGGAGCTTTAA-GATTTGACTGCCCACTGGATTTTCGGACTTATATG 1296
Db 781 AAGGAGATCAATTTGGAGCTTTAAACAATGTGACCTGCCTGCTGGATTCAGACTTGCATT 840

Qy 1297 GGGCCCGT 1304
Db 841 GGGCCCT 848

RESULT 6
BM457166
LOCUS BM457166
DEFINITION AGENCOURT_6411690 NIH_MGC_92 Homo sapiens cDNA clone IMAGE:5583427
5', mRNA sequence.
ACCESSION BM457166
VERSION BM457166.1 GI:18506206
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 898)
AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabbs-r@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: LLAM12346 row: h column: 20
High quality sequence start: 9
High quality sequence stop: 713.
Location/Qualifiers
1. .898
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5583427"
/clone_lib="NIH_MGC_92"
/lab_host="DH10B (phage-resistant)"
/lab_host="embryonal carcinoma, cell line"
/note="Organ: testis; Vector: pCMV-SPORT6; Site 1: NotI; Site 2: SalI; Cloned unidirectionally; oligo-dT primed. Average insert size 2.5 kb. Library enriched for full-length clones and constructed by Life Technologies. Note: this is a NIH_MGC Library."
BASE COUNT 225 a 228 c 254 g 191 t
ORIGIN

Query Match 34.7%; Score 477.8; DB 13; Length 898;
Best Local Similarity 77.3%; Pred. No. 6.1e-130;
Matches 669; Conservative 0; Mismatches 137; Indels 59; Gaps 5;

QY	562	TCATATGTAATGAGAGCTGAATGTTAATCTCTCAAGCAATGGGGAAAAATATCTCTGG	621
Db	35	TGCATAAGCAGCAAGAGAGCTAATGTTAATCCCAAGACCATGGGGAAATGCTCCAG	94
QY	622	ACATGTCAGAGCTCTCACAGCAAGTCCATCAATCACTTGGCTGAGGAGCTAGGAG	578
Db	95	TCATGTCAGAGACCTTCATGGCAGGCCCTCCCATCACAGGCCAGAGGCCAGAGAAA	154
QY	679	AAATGGTTTTTGGGACAGCCAGGGTCCCTGCTGTGTGTCAGCCATAGAGACTTGGTG	738
Db	155	AAGTGGTTTTTGGGCTGGGCCAGGGTCTGTGCTGTGTGTCAGCCATAGGACTTGGTG	214
QY	739	CCCTGTGTCCTAGTTAATTC	765
Db	215	CCCTGTGTCCTAGCTCTCCAGCCATGGCTGAAGAGGCAACATAGAGCTTGGGCTGGT	274
QY	766	GCTTCAGAGGTCGAAGCCCCCAAGCCTTGGCAGCTTCCAAAGTGGTGTGAGCCTTGGGT	825
Db	275	GCTTCAGAGGTCGAAGCCCCCAAGCCTTGGCAGCTTCCAAAGTGGTGTGAGCCTTGGGT	334
QY	826	GCAAGAAAGTCAAGAAATGAGGTTTTGGGAACCTCCCAATCAGATTTTCAAGAGATATATGA	885
Db	335	GCAGAGATGTCGAAGAAATGAGATTTGGGAACCTCTGCTAGATTTTCAAGATATATATGA	394
QY	886	AAACCCCTGGATGCCAGGAGAAATTTGCTGTAGGGTGGGTCCTCATGGAGAACCTCT	945
Db	395	AAACCCCTGGATGCCAGGAGAAATTTGCTGTAGGGTGGGTCCTCATGGAGAACCTCT	454
QY	946	GCAAGGTTAGTACAAAGGAAATGTTGGTGGGAGCCGCCACACAGAGTCCCGAGTGGG	1005
Db	455	CTAGGCGAGTGTGAAGGGAATG-TGGGCGAGAGCCGCCACACAGAGTCCCGTGGG	513
QY	1006	GCTCCTAGTAGAGCTGTGAGAAGAAGTCCACCATCTCC	1047
Db	514	GCAGTGCCTAGTAGAGCTGTGAGAAGAAGCCACAGTCCAGACGGTAGACCCAGAAAT	573
QY	1048	---AGACTCCAGAGGGTAGATCCACTGACAGCTTGCAGCATGTGCCTGAAAAATCCACA	1104
Db	574	GGTAGACCCAGAAATGGTAGATCAACCAACCAATTTGCTTGTGCTTGGAAAGCTACA	633
QY	1105	GACATCAGTGCACAGGCTGTGAAGCAGCAGGGAGTGTGTACCTACCAAAACCGTA	1164
Db	634	GACATCAACATCAAGCTGTGAAGCAGCAGGGAGGCTGTACTGTGCAAAAGCCACA	693
QY	1165	GTGCGAGAGCTGACCAAGACCTGGGAATCTACCTTGTGATTTGATGACCTGGACGTG	1224
Db	694	GGGCGGAGCTGACAGACCATGGAACTATCTTGTGATCAGCATGACCTGATGTG	753
QY	1225	AGACATGAGTCAAAAGAGATCAATTTGGAGCTTTAAGATTTGACTGCCCTGATTT	1284
Db	754	AGACATGAGTCAAAAGAGATCAATTTGAGCTTTAAAAATTTGACTGCCCTGATTT	813
QY	1285	CGAGCTTATATGGGCGCT-ACCCCTTGTGTGGCCAAATTTTTCATTTGGAACTGC	1343
Db	814	CAGACTTCATGAGCCCTGTAACCCCTTGTGTGGCCAAATTTTCTCTCTTGGAAACAG	873
QY	1344	CGATTTTACCAATGCCTGTACCTC 1368	
Db	874	TGATTTACCCAAATACCTGTACCC 898	
RESULT 7			
BO893011			
LOCUS	BO893011	937 bp	mRNA linear EST 16-AUG-2002
DEFINITION	AGENCOURT_8122304	Lupski_dorsal_root_ganglion	Homo sapiens cdna
ACCESSION	BO893011		
VERSION	BO893011.1	GI:22285025	
KEYWORDS			
SOURCE	human		
ORGANISM	Homo sapiens		
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		

REFERENCE	1	(bases 1 to 937)	
AUTHORS	NIH-MGC	http://mgc.nci.nih.gov/	
TITLE	National Institutes of Health, Mammalian Gene Collection (MGC)		
JOURNAL	Unpublished (1999)		
COMMENT	Contact: Robert Strausberg, Ph.D. Email: cgapbs@mail.nih.gov Tissue Procurement: Dr. James R. Lupski cdna Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Agencourt Bioscience Corporation Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov Plate: LLAM13560 row: c column: 16 High quality sequence stop: 722.		
FEATURES	Location/Qualifiers		
source	1..937		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
	/clone="IMAGE:6179271"		
	/clone_lib="Lupski_dorsal_root_ganglion"		
	/sex="male"		
	/tissue_type="dorsal root ganglia"		
	/dev_stage="adult, 36 yr"		
	/lab_host="DH10B"		
	Note: Vector: PCMV-SPORT6 (Life Technologies); Site_1: NotI; Site_2: SalI; cdna made by oligo-dT priming. Directionally cloned using the following adaptors: 5'-TCGACCCAGGCTCG-3' and 5'-GACTAGTTCTAGATCGGAGCGCCCT(15)-3'. Size selected > 1 kb for average insert length 1.7 kb. This is a primary library, non-amplified. Library constructed by Life Technologies and donated by J. Lupski, M.D./Ph.D. (Baylor College of Medicine) and is available through Life Technologies.		
BASE COUNT	235 a	226 c	258 g
ORIGIN		214 t	4 others
	Query Match	34.1%	Score 469.8; DB 14; Length 937;
	Best Local Similarity	75.4%	Pred: No. 1.4e-127;
	Matches 682; Conservative	0; Mismatches 149; Indels	74; Gaps 5;
QY	514	AAGAAAAACCCATTTCTTCTGAGGGAAATTTCAAGCTGGCTGCAGAAATTTGCATATGTAAT	573
Db	3	ATGAAATTTGCAATTTCTTGAATAGAAATTTCAAGCCAGCTGCAGAAATTTGCATAAGGAAG	62
QY	574	GAGGAGCTGAATGTTAATCCTCAAGACAATGGGAAAATATCTCTGGACATGTGAGAGG	633
Db	63	GAGGAGCAAAATGTTAATCCCAAGACAATGGGAAAATGTCTCCAGGCCATGTGAGAGG	122
QY	634	TCCTTCACAGCAGTCCATCAAAATCACTGGCTCGAGGCCCTAGGAG---AAAATGGTTTGT	690
Db	123	CCTTCAGGCGAGTCCCTCCATCAGGCCCAAGAGCCCTAGGAGGAAAAGTGGTTTCAT	182
QY	691	GGGACAGGCCCAAGGTCCTGTGTGTGTCAGCCCTAGAGACTTGTGCTGTGTGTCCTCA	750
Db	183	GGGCTGGGCCCAAGGTCCTCAAGCTATGTGACGCTTGGGTTTGTGCTCACTGCATCCCA	241
QY	751	GTTAATTCAGCTG-----TGCTTCAGAGGT 777	
Db	242	GCTGCCCAAGCTGTGGCTGAAAGGGGACAAATAGAGCTCGGGGCCCTGTGCTTTAGATGT	301
QY	778	GCAAGCCCCCAAGCCTTG-----GCAGCTTCCCAAGTGGTGTGAGCCTGTGGGTGCAAG	831
Db	302	GCAAGCCCCATGCCCTGTGGTGTGAGCTTCCATGTGGTGTGAGCCTCGAGGTGCACAG	361
QY	832	AAGTCAAGAAATTTAGGTTTGGGAACCTCCAATCAGATTTTCAAGATATATATGAAACCC	891
Db	362	AAGTCAAGAAATTTAGGTTTGGGAACCTCCACCTAGATTTCAAGATATGATGAAATGCC	421
QY	892	TGGATGCCCGCAGCAAGATTTTGTCTAGGGTGGGTCTCTCATGAGAACCTCTGCAAGG	951
Db	422	TGGATGCCCGCAGCAAGATTTTGTCTGCAAGGGTGGGCTCTCATGAGAACCTCTGCTAGG	481

ACCESSION BQ723415
VERSION BQ723415.1 GI:21862312
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 941)
AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Dr. James R. Lupski
cDNA Library Preparation: Life Technologies, Inc.
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: LLAM13596 row: b column: 17
High quality sequence stop: 606.
Location/Qualifiers
1..941
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:6193072"
/clone_lib="Lupski_sympathetic_trunk"
/sex="male"
/tissue_type="sympathetic trunk"
/dev_stage="adult, 16 yr"
/lab_host="DH10B"
/note="Vector: pCMV-SPORT6 (Life Technologies); site_1:
NotI; site_2: SalI; cDNA made by oligo-dT priming.
Directionally cloned using the following adaptors:
5'-TCGACCCACGGCTCCG-3' and
5'-GACTAGTTCTAGATCGGAGCGGCCCT(15)-3'. Size selected >
1 kb for average insert length 1.9 kb. This is a primary
library, non-amplified. Library constructed by Life
Technologies and donated by J. Lupski, M.D./Ph.D. (Baylor
College of Medicine); available through Life
Technologies."

BASE COUNT 257 a 208 c 267 g 209 t

Query Match 32.8%; Score 451.4; DB 14; Length 941;
Best Local Similarity 78.0%; Pred. No. 4.1e-122;
Matches 672; Conservative 0; Mismatches 146; Indels 43; Gaps 9;

QY 391 GAAGAAATTTCTAAGCAGCAAGCATTTCAAGAGTGACTTGGTGCTGTTAAAGGCATTTC 450
DB 51 GAAGAAATTTCTAAGCAGCAAGCATTTCAAGAGTGCTCTTGGTGCTGTTAAAGCATTC 110
QY 451 AGTTTCATAAGGAGCAGAGCATTAAGAGTTTCAGAAAAATTTGCACCCCTGACAAATGTGATA 510
DB 111 AGTTTAAAAGGGAACAGACGATAAAAGTTTGGAAAAATTTGCACCTGCAACAATGCAATA 170
QY 511 AAAAAGAAAAACCCATTTTCTCAGG-GGAATTCAGCTGGCTGCAGAAATTTGCATATG 569
DB 171 GAAAGAAAAATCCCATTTTCTCAGGAGAAAGTCAAGCTGGCTGCAGAAATTTACATAAG 230
QY 570 TAATGAGGAGCTGAATGTTTAATCCTCAAGACAAT-GGGGAAAAATATCTCCTGGACATGTC 628
DB 231 TAACAAGAGCCCAATGTTAATCCCCAGACAATGGGGGTAAATGTCTCCAGAGCATATC 290
QY 629 AGAGGTCTTCACAGCAGTCATCAATCACTGGCCCTGGAGGCGCTAGGAG---AAATGGT 685
DB 291 AGAGGTCTTCACGGCAGCCCTCCCTTTACAGGCGCTGGGGCGCTAGGAGGAAAAAGTGT 350
QY 686 TTTGTGGGACAGCCCGAGGTCCTGCTGTGTGTCAGCCTAGACACTTGGTGCCTGTG 745
DB 351 TTCGTGGGCGCTGACACAGAGGGTCCCCATGCTGTGTGCAGCCTAGGAGTCTGGTGCCTGCA 410

QY 746 TCCAGTTAAATTC-----GCTGTGGCTTCAG 772
DB 411 TCCAGCTGCTCCAGCTGTGGCTGAAAGGGGCTATCACAGAGCTCAGGCTGTGGCTTCAG 470
QY 773 AGGTGTCAGCCCAAGCCTTGGCAGCTTCCAAGTGTGTTGAGCCTGTGGGTGCAAGA 832
DB 471 AGGTGTCAGTCCCAAGCCTTGGCAGCTTCCAAGTGTGTTGAGCCTGCAAGTACATGA 530
QY 833 AGTCAAGAAATTGAGTGTGGGAACTCCAAATCAGATTTTCAGAAAGATATATGAAACCCCT 892
DB 531 AGTCAAGAAATTGAGTGTGGGAACTCCCTAGATTTTTCAGAGGTACGGAATACCT 590
QY 893 GGATCCCGCAGCAGAAAGTTTGTAGGGTGGGTCTCTATGAGAGAACTCTGCAAGGG 952
DB 591 GGATCCCGCAGAGAAAGTTTGTGCGAGCAGCAGGCGCTCATGAGAACTTCTGCTAG 650
QY 953 TAGTACAAAAGGAAATGTTGGTGGGAGCCCGCCACACAGAGTCCCAGTGGGGCTC-CA 1011
DB 651 TAGTGGGAGGAAATGTAGGTTCAGAGCCCGCCACACAGAGTCCCTACTGAGACACTGC 710
QY 1012 TCTAGTAGAGCTGTGAGAAAGTCCACCATCC-TCCAGACTCCAGAGGGGTAGATCCAC 1070
DB 711 CCTAGTGAGCTGTGAGAAAGGCCATCATCTTTTCAGACCCAGGAATGGTAGATCCAC 770
QY 1071 TGACAG-CTTGACGATGTGCTGAAATCCACAGACTCAGTCCAGCCTGTGAAAG 1129
DB 771 TGACAGCCTTGACCTGTGTGCTGGAAAGGCCACAGACTCAACGCCAGCCCATTAAG 830
QY 1130 CA-GCAGGATGGAGTGTACCTTACAAAACCGTAGTGGCAGAG-CTGACCAAGACCGT 1187
DB 831 CACCAGGAGGAGGCTGTACCTTGCAAAGCCACAGGGGTGGGACAGCCCAAGACTAT 890
QY 1188 GGAATCTACCTCTTGCAATTG 1208
DB 891 GGGACCCACCCTCTTTTTTG 911

RESULT 10
BQ423563
LOCUS AGENCOURT_7845919 NIH_MGC_72 Homo sapiens cDNA clone IMAGE:6064274
DEFINITION 5', mRNA sequence.
ACCESSION BQ423563
VERSION BQ423563.1 GI:21118878
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 876)
AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: ATCC/DCTD/DTF
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: LLAM13338 row: 1 column: 03
High quality sequence stop: 674.
Location/Qualifiers
1..876
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:6064274"
/clone_lib="NIH_MGC_72"
/tissue_type="melanotic melanoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: skin; Vector: pCMV-SPORT6; Site_1: NotI;

FEATURES
source

Slte_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 2 kb. Library constructed by Life
Technologies."

```
BASE COUNT      242 a      246 g      187 t      1 others
ORIGIN
Query Match      31.6%; Score 435.6; DB 14; Length 876;
Best Local Similarity 78.9%; Pred. No. 1.9e-117;
Matches 610; Conservative 0; Mismatches 120; Indels 43; Gaps 6;
Qy 392 AAGAAATTTCTAAGCAGCAAGCATTTCAAGAGGTGACTTGGTGTCTGTTAAAGGCATTCA 451
Db 101 AAGAAATGTCAGAACCAAGCAAGCATTTCAAGAGGTGACTTGGTGTCTGTTAAAGGCATTCA 160
Qy 452 GTTTTCATAAGGAGGAGCAGCATGAAGAGTTCAGAAAAATTTGCACCCCTGCACAAATGTGATAA 511
Db 161 GTTTTAAAGGGAACAGACACCTTAAGAGTTCAGAAAAATTTGCAGCCCTGACTATGTGATAG 220
Qy 512 AAAAGAAAAACCCATTTTCTGAGGGGAAATTCAGAGCTGGCTGGCAGAAAATTTGCATATGTA 571
Db 221 AAAAGAAAAACCCATTTTCTGGGAGAAATTCAGGCCAGCTGCAGAAATTTGCATAAGTA 280
Qy 572 ATGAGGAGCTCAATGTTAATCTCAAGACAATGGGAAATATCTCCTGGACATGTCCAGA 631
Db 281 GTAAGGAGCTTAATGTTAATCTCCCAAGACCACTGGAAATATGTTCCAGGTCAATGTCCAGA 340
Qy 632 GGTCTTCACAGAGTCCATCAATCACTGGCTGGAGGCC---TAGGAGAAAAATGGTGTTC 688
Db 341 GACCTTCACGCGCAGCCCAACCATCACAGGCTGGAGGCCAGGAGGAGAAATGGTGTTC 400
Qy 689 GTGGCAGGCCAGGCTCCCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 748
Db 401 ATGGGCCAGGCCAGGCTCCCTGTATTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 460
Qy 749 CAGTTAATTCA-----GCTGTGCTTCAGAGG 775
Db 461 CAGCTTCCCAAGCTGTAGCTAAAGGGGCCAATGTACAGCTCGGGCTGTGCTTCAGAGG 520
Qy 776 GTCAAGCCCCAAGCCTTTGGCAGCTTCCAAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 835
Db 521 GTGGAAGCCCCAAGCCTTGGCAGCTTCCATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 580
Qy 836 CAAGAATTGAGGTTTGGAACTTCAATCAGATTTTCAGAAATATATGGAACCCCTGGA 895
Db 581 CAAGAATTGAGGTTTGGAACTTCAATCAGATTTTCAGAAATATATGGAATGCTTGA 640
Qy 896 TGCCAGGCAGAAATTTGCTGTAGGGTGGGTCTCTCATGAGAACCTTCGCAAGGGTAG 955
Db 641 TGCCAGGCAGAAATTTGCTGTAGGGTGGGACCTTCATGAGAGCCCTCTCTAGGCCAG 700
Qy 956 T-ACAAAAGGAAATGTTGGGTGGAGGCCCCACACAGAGTCCCGAGTGGGCG---TGCA 1011
Db 701 TGTGAAAGGAAATGTTGGCGTTTGAACCCCAATGAAGTCCCTAGTGGGCACTCCCC 760
Qy 1012 TCTAGTAGAGCTGTGAGAAAGTCCACCATTCTCCAGACTCCAG-AAGGGTAGATCCAC 1070
Db 761 TAGTGGAGCTTGTGAAAAAGAGGGCCACTGTCTCCAGACCCCAAAATGTTAGATCCAC 820
Qy 1071 TGACAGCTTGCAGCATG--TGCCCTGAAAAATCCACAGACACTCAGTGCACGCC 1121
Db 821 TGACAGCTTGCACCTGTGGCACTCGGAAAAAACCCGACAGACACTCCATGCCAGCC 873
```

RESULT 11
AG086951
LOCUS AG086951 717 bp DNA linear GSS 03-NOV-2001
DEFINITION Pan troglodytes DNA, clone: PTB-085H08.F, genomic survey sequence.
ACCESSION AG086951
VERSION AG086951.1 GI:16638753
KEYWORDS GSS.
SOURCE Pan troglodytes male lymphoblast DNA, clone_lib:PTB Chimpazee Male
BAC Library clone:PTB-085H08.F.
ORGANISM Pan troglodytes

Eukaryota; Metazoa; Chordata; Cranlata; Vertebrata; Euteleostomi;
Mammalia; Eutherla; Primates; Catarrhini; Hominidae; Pan.
1
Fujlyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
Totoki,Y., Watanabe,H. and Sakaki,Y.
BAC end Sequences of Library PTB
Unpublished
2 (bases 1 to 717)
Fujlyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
Totoki,Y., Watanabe,H. and Sakaki,Y.
Direct Submission
Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
1-7-22 Suehiro-chou,Tsrumi-ku, Yokohama, Kanagawa 230-0045, Japan
(E-mail:chimbesc@sc.riken.go.jp, URL:http://hgp-gsc.riken.go.jp/,
Tel:81-45-503-9111, Fax:81-45-503-9170)
Clones are derived from the chimpanzee BAC library PTB. This BAC end
was generated during the R&D process and may have higher chance of
clone tracking errors.

PRIMERS
Sequencing: -21M13

LIBRARY
Vector : pKS145
R.Site 1 : SacI
R.Site 2 : SacI
1. 717
Location/Qualifiers
/organism="Pan troglodytes"
/db_xref="taxon:9598"
/clone="PTB-085H08.F"
/sex="male"
/cell_type="lymphoblast"
/clone_lib="PTB Chimpanzee Male BAC Library"
BASE COUNT 175 a 189 c 195 g 158 t
ORIGIN

Query Match 31.3%; Score 430.4; DB 17; Length 717;
Best Local Similarity 82.4%; Pred. No. 6.1e-116;
Matches 506; Conservative 0; Mismatches 106; Indels 2; Gaps 1;
Qy 745 GTCCAGTTAATTCAGCTGTGCTTCACAGGTGTCAGACCCCAAGCCCTTGGCAGCTTCCA 804
Db 90 GTCCCGGAGCTCAGGCTGTGCTTCAGAGGTGTCAGACCCCAAGCCCTTGGCAGCTTCCA 149
Qy 805 AGTGTGTGTAGCCCTGTGGGTGCAAGAAAGTCAAGAAATTTGGAGCTTGGGAACCTCCAATC 864
Db 150 CATGTGTAAAGCTGTGAGTGCACAGAAATCAAGAACTGGGTTTGGGAACCTCTGCTC 209
Qy 865 AGATTTCAAGAGATATATGGAAACCCCTGGATGCCAGGCAGAGAACTTTGCTGTAGGGTG 924
Db 210 AGATTTCAAGAGATATATGGTAATACCTGGATGCCAGGCAGAGAAATTTGCTGTAGGGAG 269
Qy 925 GGGTCTCATGAGAACCTCTGCAAGGGTAGTACAAAAGGGAATTTGGGTGGGAGGCC 984
Db 270 GGGGCTCATGAAAACCTCTGCTAGGGCAGTGTGGGAGAGAAATGTGGGTTCAGAGCCA 329
Qy 985 CCACACAGAGTCCCGAGTGGGCTCCATCTAGTAGAGCTGTGAGAAAGTCCACCATCC 1044
Db 330 TCACACAGTCCCTTACTTGGGACCCGCTAGTGAGCTGTGAAAAGTGGGCCACTATTC 389
Qy 1045 TCCAGACTCCAGAGGGTAGATCCACTGACAGCTTTCAGAGCATGTGCTGAAAATTCACA 1104
Db 390 TCCAGACCCAGAAATGTTAGATCCACTGACAGCTTTCACCATGCATGCAAGAACCCACA 449
Qy 1105 GACACTCAGTCCCGCTGTGAAAGCAGCAGGATGGAGTCTGTACCTACAAAACCGTA 1164
Db 450 GACACTCAATGCCAGCCCTGTGAAAGCAGCTGGGAGGAGACTGTACCATGCAAGACCCACA 509
Qy 1165 GTGGCAGAGCTGACCAAGACCGCTGGGAATCTACCTCTTGTGCAATTTGCATGACCTGACCGTG 1224
Db 510 GGGCAGAGCTGCCCAAGACCATGGGAACCCACCTCTTGCATCATCATGACCTGGATGTG 569
Qy 1225 AGACATGGAGTCAAAAGAGATCATTTTGGAGCTTTAAGATTGTGACTGCCCCACTGGATTT 1284


```
FEATURES
Source      Location/Qualifiers
1. .676
/organism="Pan troglodytes"
/db_xref="taxon:9598"
/clone="PTB-053H24.R"
/sex="male"
/cell_type="lymphoblast"
/clone_lib="PTB Chimpanzee Male BAC Library"
BASE COUNT 164 a 165 c 191 g 156 t
ORIGIN

Query Match      30.5%; Score 419.8; DB 17; Length 676;
Best Local Similarity 82.3%; Pred. No. 8.2e-113;
Matches 506; Conservative 0; Mismatches 107; Indels 2; Gaps 2;

QY 759 TGTGCTGCTCAGAGGGTCAAGAGCCCAAGCCTTGGCAGCTTCCAAGTGGTGTGAGCC 818
DB 62 AGCATGGCTTCTAAGGTGCAAGCCTCAAGCCTTGGCAGCTTCCAAGTGGTGTGAGCT 121
QY 819 TGTGGGTGCAAGAAGTCAAGAAATTGAGGTTTGGGAACCTCCAATCAGATTTTCAGAGAT 878
DB 122 GGCAAGTGCACAGAAGTCAAGAACTGGGGTTTAGGAACCTCTACCTAGATTTTCAGAGAT 181
QY 879 ATATGGAACCCCTGGATGCCAGGCAAGAGTTTGTGTAGGGGTGGGGTCTTCATCGAG 938
DB 182 GTATGGAATGCCCTGGATGTCAGGCAAGAGTTTGGCGAGGACGGGGTCTCTTGGAG 241
QY 939 AACTCTGCAAGGTAGTACAAAAGGGAATTTGGGTGGGAGCCGCCACACAGAGTCCC 998
DB 242 AACTCTGCTAGGCAATGTGGAAGGGAATTTGGGGTTGGAGTCCCGCCAGAGTCCC 301
QY 999 CAGTGGGGCTCCATCTAGTAGAGCTGTGAGAAGAAGTCCACCATCTCCAGACTCCAGAA 1058
DB 302 TACTGGGGCACCATCTAGTGGAGCTGTGGAAG-AGGGCACTGACCTCCAGACCCAGAA 360
QY 1059 GGTAGATCAGCTGACAGCTTGAGCATGTGCTGAGGAGTGGAGTCCAGACACTAGTCCCA 1118
DB 361 TGTGTAGTCCACTGACAGCTGTACACCATGTGCTGAAAGAGCTGTAAACACTCAACACCA 420
QY 1119 GCCTGTGAAAGCAGCAGGAGTGGAGTCTGTACCTTACAAACCGTAGTGCAGAGCTGAC 1178
DB 421 GCCCATGAAGCATCTGGGAGGAGGTTGTACCTTGCAGATCCACAGGGGGGAGCTGCC 480
QY 1179 CAAGACCGTGGGATCTACCTCTTGTGATGTGATGACCTGGAGCTGAGACATGAGAGTCAA 1238
DB 481 CAAGACCATGGAGCCACCTCTTGCATCAGTGTGACCTGGATGTGAGACATGGAGTCAA 540
QY 1239 AAGAGATCATTTTGGAGCTTTAAGATTGACTGCCCCACTGGATTTTCGGACTTATATGGG 1298
DB 541 AGGAGATCATTTTGTAGTTTAAAGATTGACTGCCTCGCTGGATTTTGGACTAGCATGGG 600
QY 1299 GCCCGTA-CCCTTTGTTTGGCAATTTTTCATTTGGAAGTCCGCTATTTTACCCAAT 1357
DB 601 GCCTGAGCCCCCTTGTGTTTGGCAATTTCTCCCAATTTGGAATGGCTGTATTTACCCAAT 660
QY 1358 GCCTGTACCTCCCAT 1372
DB 661 GCCTGTACCCCAT 675

RESULT 14
LOCUS      BM926985
DEFINITION AGENCOURT_5681877 NIH_MGC_121 Homo sapiens cDNA clone IMAGE:5767880
5', mRNA sequence.
ACCESSION BM926985
VERSION    BM926985.1 GI:19377364
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 1063)
```

NIH-MGC <http://mgc.nci.nih.gov/>.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgaps-r@mail.nih.gov
Tissue Procurement: Life Technologies, Inc.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: LLAMI2826 row: n column: 09
High quality sequence stop: 613.

FEATURES
source
1. .1063
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5767880"
/clone_lib="NIH_MGC_121"
/lab_host="DH10B"
/note="Organ: brain; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: EcoRV (destroyed); RNA source anonymous pool of 3
fetal brains, female age 20 weeks, female age 24 weeks,
and male age 26 weeks. Library is oligo-dT primed and
directionally cloned (EcoRV site is destroyed upon
cloning). Average insert size 1.7 kb, insert size range
0.7-3.5 kb. Library is normalized and enriched for
full-length clones and was constructed by C. Gruber
(Invitrogen). Research Genetics tracking code 017. Note:
this is a NIH_MGC Library."

BASE COUNT 231 a 311 c 244 g 277 t
ORIGIN

Query Match 30.2%; Score 416.4; DB 14; Length 1063;
Best Local Similarity 80.9%; Pred. No. 9.8e-112;
Matches 497; Conservative 0; Mismatches 116; Indels 1; Gaps 1;

QY 764 TGGCTTCAGAGGGTGCAGGCCCAAGCCTTGGCAGCTTCCAAGTGGTGTGAGCTGTGG 823
DB 764 TGGCTTCAGAGGGTGCAGGCCCAAGCCTTGGCAGCTTCCAAGTGGTGTGAGCTGTGG 128
QY 824 GTGCAAGAAGTCAAGAAATTGAGGTTTGGGAACCTCCCAATCAGATTTTCAGAGATATATG 883
DB 129 GTGCACAGAAGTCAAGAGTTGAGGTTTGGGAATCTCTGCTGGATTTTCAGAGGATGTATG 188
QY 884 GAACCCCTGGATGCCAGGCGAAGTTTGTGTAGGGTGGGGTCTCATGGAGAACT 943
DB 189 GGAACATCTAGATGCCAGGCAAACTTGTGTCAGGGGAGGGGTCTTCTTTGGAGAACT 248
QY 944 CTGCAAGGGTAGTACAAAGGGAATGTGGTGGGAGCCCCACACAGAGTCCCCAGTG 1003
DB 249 CTGTTAGGCGATGCCAAAGAAATGTGGGTGGAGCCCCCAACAAAGATATCTACTG 308
QY 1004 GGGCTCCATCTAGTAGAGCTGTGAGAAGAAGTCCACCATCTCCAGACTCCCAAGAGGTA 1063
DB 309 GGGCACCCCTAGTGGAGCTGTGAGAAGAGGGTCAACCATCTCCCAACCCCAATGTA 368
QY 1064 GATCCACTGACAGCTTGCAGCATGTGCTGAAAATTCACAGACTCAGTGCAGGCTG 1123
DB 369 GAGCTACTGACAGCTTGCACCATTCACCTGGAAGAACCCACAGATCTCAACGCGACACA 428
QY 1124 TGAAGCAGCAGGATGAGTCTGTACCTACAAACAGCTAGTGGCAGAGCTGACCAAGA 1183
DB 429 TGAAGCAGATGGGAGAGAGTTGTACCTGCAATGCCAGGGGAGAGCTGCCCCAAGA 488
QY 1184 CCCTGGGAATCTACCTCTTGCATTTGTATGACCTGGAGCTGAGACATGAGGTGAGAGAG 1243
DB 489 CCATAGAACCCACTCTTGCATCAGCATGACCTGTGATGAGACCTGGAATCAAGAGAG 548
QY 1244 ATCATTTTGGAGCTTTAAGATTGACTGCCCACTGGGATTTTCGAGCTTATATGGGGCCG 1303
DB 549 ATCATTTTGGAGCTTGAATAATTTGACTGCTGTATAGACTTTCGGGCTTTCGCTGCTGCTG 608

```
QY 1304 -TACCCCTTTGTTGGCAATTTTTCATTGGAATCGCGTATTTACCAATGCCCTG 1362
|||||
Db 609 CAACCCCTTTGTTGGCAATTTTCTCCATTGGAATGGCTGTATTACCCCAATACCTG 668

QY 1363 TACCTCCATTGTAT 1376
|||||
Db 669 TACCCCCCATGTAT 682

RESULT 15
BQ422247/c
LOCUS BQ422247 1080 bp mRNA linear EST 23-MAY-2002
DEFINITION AGENCOURT_7766790 NIH_MGC_92 Homo sapiens cDNA clone IMAGE:6015590
5', mRNA sequence.
ACCESSION BQ422247
VERSION BQ422247.1 GI:21117562
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 1080)
NIH-MGC http://mgi.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: gcgabs-r@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM13211 row: o column: 15
High quality sequence stop: 702.
FEATURES
source
1. 1080
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:6015590"
/clone_lib="NIH_MGC_92"
/tissue_type="embryonal carcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: testis; Vector: PCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally; oligo-dt primed.
Average insert size 2.5 kb. Library enriched for
full-length clones and constructed by Life Technologies.
Note: this is a NIH_MGC Library."
BASE COUNT 239 a 304 c 221 g 316 t
ORIGIN
```

```
Query Match 30.1%; Score 414.8; DB 14; Length 1080;
Best Local Similarity 79.9%; Pred. No. 2.9e-111;
Matches 566; Conservative 0; Mismatches 102; Indels 40; Gaps 5;

QY 350 GGAAGAAATTTCTAGACGACCAAGCAATTCACAGAGTGAGTGGGTGCTGTTAAAGGCATT 449
|||||
Db 708 GGAAGAAATTTCTAGACGACCAAGCAATTCACAGAGTGAGTGGGTGCTGTTAAAGGCATT 649
|||||
QY 450 CAGTTTTCATAGGGAGGACGACATAGAGTTACAGAAATTTGCACCTTGACAATGTGAT 509
|||||
Db 648 CAGTTTTCATAGGGAGGACGACATAGAGTTACAGAAATTTGCACCTTGACAATGTGAT 589
|||||
QY 510 AAAAAAGAAAAA-CCATTCTTGTAGGGGAAATTCACAGTGCCTGCAGAAAAATTTGCATAT 568
|||||
Db 588 AGAAAAAGAAAAATCTCATTTCTGAGGAGAAATGGAAGCAGGCTGCAGAAATTCGCCTAA 529
|||||
QY 569 GTAATGAGGAGCTGATGTTAATCTCAAGACAAATGGGGAAAAATCTCCTGGACATGTC 628
|||||
Db 528 GTAATGAGGAGCAAAATGTTAATCTCAAGACAAATGGGGAAAAATGCTCCAGGGTATGTC 469
|||||
```

Search completed: February 20, 2003, 06:15:08
Job time : 3510.23 secs

```
QY 629 AGAGGTCTTTCACGACAGTCCATCAATCACTGGCCTGGAGCCT---AGGAGAAAAATGGT 685
|||||
Db 468 AGA-GTTTTTCATGGCAGCCCTTCCATCACAGGCCAGAGGCCCTACAAGAAAAATGGT 410
|||||
QY 686 TTTGTTGGGACAGGCCCAGGGT--CCCTGTGCTGTGTGTCAGCCCTAGAGACTTGGTGCCCTG 743
|||||
Db 409 TTTCTGGGCCAGGCCCAGGGTACACTCGTGTGTGTGTCAGCCAAAGAGACTTGGTGCTCTG 350
|||||
QY 744 TGTCCCAAGTTAATTCA-----GCTGTGCTTTC 770
|||||
Db 349 CATCCCAAGTCGCTCCAGACATGGCTGAAAAGGGGCCAACATAGAGTTTCGGGCAAGTGGCTTC 290
|||||
QY 771 AGAGGTGTCAAGCCCAAGCCCTTGGCAGCTTCCAACTGGTGTGTGAGCCTGTGGTGTGCAAA 830
|||||
Db 289 ACAGGTGTCAAGCCCAAGCCCTTGGCAGCTTCCATGTGTGTGAGCCCTGAGAGTGCACA 230
|||||
QY 831 GAAGTCAAGAATTGAGTTTGGGAACCTCCAATCAAGTTTCAGAAAGATATATGGAACCC 890
|||||
Db 229 AAAGTCAAGAACTGGGGTTTGGGAACCTCTGCCTAGATTCTGAAGATGTATGGAATGC 170
|||||
QY 891 CTGGATGCCAGGCACAGAAAGTTTCTGTAGGGTGGGGTCTCTCATGGAGAACCTCTGCAAG 950
|||||
Db 169 CTGGATGCTCAGGCATAAAGTTTGTCTAGGGTGGGGTCTCTCATGGCAACCACTTTAG 110
|||||
QY 951 GGTAGTACAAAAAGGGGAAATGTTGGGTGGGAGCCGCCACACAGAGTCCCGAGTGGGGCTCC 1010
|||||
Db 109 GGCAGTGGAGAGGAATAATGTGAGGTACAGAGCCGCCACACAGAGTTCCCGCTGGGGCACC 50
|||||
QY 1011 ATCTAGTAGAGCTGTGAGAAAGTCCACCATCTCTCCAGACTCCAGAA 1058
|||||
Db 49 ACCTAGTGGAGCTGTGAGAAAGAGGGCCACTGTCTCTCCAGACCTCAGAA 2
|||||
```

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 19, 2003, 18:01:21 ; Search time 110.805 Seconds
(without alignments)
6329.459 Million cell updates/sec

Title: us-09-997-610-1_copy_2_1378

Perfect score: 1377

Sequence: 1 atagtggtacatacctgtctt.....gcctgtacctccattgtatg 1377

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 424239 seqs, 254661826 residues

Total number of hits satisfying chosen parameters: 848478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published_Applications_NA:*

- 1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
- 2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
- 3: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
- 4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
- 5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
- 6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
- 7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
- 8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
- 9: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
- 10: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
- 11: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
- 12: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
- 13: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
- 14: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1377	100.0	1381	9	US-09-997-610-1
2	1377	100.0	1731	9	US-09-997-610-5
3	988.4	71.8	1377	9	US-09-997-610-3
4	988.4	71.8	1731	9	US-09-997-610-7
5	641.6	46.6	155074	9	US-10-026-188-6
6	618.2	44.9	465237	10	US-09-933-267A-1
7	610.2	44.3	180557	12	US-10-003-806-6
8	610.2	44.3	180557	12	US-10-003-806-9
9	603.4	43.8	170834	10	US-09-835-233-7
10	590.4	42.9	202001	10	US-09-734-674-3
11	553.8	40.2	684973	10	US-09-263-959-1
12	545.2	39.6	69327	10	US-09-777-921A-3
13	523.6	38.0	25603	9	US-09-819-607-3
14	523	38.0	1946	10	US-09-864-761-2925
15	502.8	36.5	148567	9	US-10-254-869-3
16	502.8	36.5	148567	9	US-09-801-876B-3
17	469.6	34.1	180557	12	US-10-003-806-6
18	469.6	34.1	180557	12	US-10-003-806-9
19	463	33.6	197997	10	US-09-822-246-3

c	20	450.8	32.7	143068	10	US-09-967-768A-316	Sequence 316, App
c	21	439.6	31.9	49136	10	US-09-768-877-1	Sequence 1, Appli
c	22	434.8	31.6	11469	10	US-09-764-877-2791	Sequence 2791, Ap
c	23	430.8	31.3	761	10	US-09-864-761-19706	Sequence 19706, A
c	24	420.2	30.5	15819	10	US-09-764-877-2765	Sequence 2765, Ap
c	25	417	30.3	12104	10	US-09-764-877-2761	Sequence 2761, Ap
..	26	400.2	29.1	1666	9	US-10-001-835-76	Sequence 76, Appl
	27	398.2	28.9	55795	10	US-09-880-107-1543	Sequence 1543, Ap
c	28	392	28.5	577	10	US-09-864-761-6870	Sequence 6870, Ap
c	29	391.6	28.4	3816	10	US-09-880-107-2298	Sequence 2298, Ap
c	30	391.2	28.4	7642	10	US-09-764-877-2764	Sequence 2764, Ap
c	31	382	27.7	11071	10	US-09-764-877-2793	Sequence 2793, Ap
c	32	370.2	26.9	42999	10	US-09-740-029-3	Sequence 3, Appli
c	33	369.4	26.8	63588	9	US-10-243-735-3	Sequence 19812, A
c	34	362	26.3	3693	10	US-09-864-761-19812	Sequence 3941, Ap
c	35	356.6	25.9	2093	10	US-09-880-107-3941	Sequence 7, Appli
c	36	356.4	25.9	170834	10	US-09-835-232-7	Sequence 6964, Ap
c	37	347.4	25.2	594	10	US-09-864-761-6964	Sequence 9752, Ap
c	38	341.2	24.8	573	10	US-09-864-761-9752	Sequence 623, App
c	39	327	23.7	847	10	US-09-764-887-623	Sequence 8786, Ap
c	40	311.6	22.6	575	10	US-09-864-761-8786	Sequence 5584, Ap
c	41	311.4	22.6	489	10	US-09-864-761-5584	Sequence 9187, Ap
c	42	304.8	22.1	578	10	US-09-864-761-9187	Sequence 849, App
c	43	298.2	21.7	465	10	US-09-864-761-649	Sequence 8370, Ap
c	44	292.8	21.3	597	10	US-09-864-761-8370	Sequence 819, App
c	45	290	21.1	471	10	US-09-864-761-819	

ALIGNMENTS

RESULT 1

US-09-997-610-1
; Sequence 1, Application US/09997610
; Patent No. US20020156244A1
; GENERAL INFORMATION:
; APPLICANT: Fox, Brian
; TITLE OF INVENTION: ADIPOCYTE COMPLIMENT RELATED PROTEIN
; TITLE OF INVENTION: ZACRP13
; FILE REFERENCE: 00-96
; CURRENT APPLICATION NUMBER: US/09/997.610
; CURRENT FILING DATE: 2001-11-29
; PRIOR APPLICATION NUMBER: US 60/253,924
; PRIOR FILING DATE: 2000-11-29
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 1381
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (2)...(1381)
US-09-997-610-1

Query Match	100.0%	Score 1377;	DB 9;	Length 1381;
Best Local Similarity	100.0%	Pred. No. 0;		
Matches 1377;	Conservative	0;	Mismatches	0;
Indels	0;	Gaps	0;	
Qy	1	ATAGTGGTACATACCTGTCTTTAATAACGCAGTCATTTAGCATGTAGAAAGTTGCTGGACCT	60	
Db	2	ATAGTGGTACATACCTGTCTTTAATAACGCAGTCATTTAGCATGTAGAAAGTTGCTGGACCT	61	
Qy	61	CCAGCACACCCAGCCGCCAGAGAGAGTGGGGCTCTCTGTGCACAGGTTTACCACAA	120	
Db	62	CCAGCACACCCAGCCGCCAGAGAGTGGGGCTCTCTGTGCACAGGTTTACCACAA	121	
Qy	121	TATACAGAGAAATAAGTGAATGACAAATGCCCTCTCTCTGATATAGAAGGTCAGCC	180	
Db	122	TATACAGAGAAATAAGTGAATGACAAATGCCCTCTCTCTGATATAGAAGGTCAGCC	181	
Qy	181	TTTACTGTGAAGCTCAGTGGAAAACCTTCCTCTTCTTCAAGCCCATCATCTTCACAGG	240	

Db 182 TTTACTGTGAAGCTCAGTGGAAACCTTCCTCTCCCTTTCAAGCCCATCATCTTTCACAGG 241
QY 241 GTCCTGTACAATGCCAGAGGATTTAAAGAGGCGCATGGGAGTCTTTCTCTTGCGAGGGTG 300
Db 242 GTCCTGTACAATGCCAGAGGATTTAAAGAGGCGCATGGGAGTCTTTCTCTTGCGAGGGTG 301
QY 301 CTGGGAATTTACTACTCCAGCTTTGATGTGTAGCTGCATCATTTGCAAGGTGAATATTTGG 360
Db 302 CTGGGAATTTACTACTCCAGCTTTGATGTGTAGCTGCATCATTTGCAAGGTGAATATTTGG 361
QY 361 CTAATGAGGAACCAATTTTGCTTAATAGGAAGAAATTTCTAAGCAGCAAAAGCATTTCAA 420
Db 362 CTAATGAGGAACCAATTTTGCTTAATAGGAAGAAATTTCTAAGCAGCAAAAGCATTTCAA 421
QY 421 GAGGTGACTTGGTGTGTTAAAGGCATTCAGTTTCATAAGGAGCGCAGAGCATAGAGT 480
Db 422 GAGGTGACTTGGTGTGTTAAAGGCATTCAGTTTCATAAGGAGCGCAGAGCATAGAGT 481
QY 481 TCAGAAAATTTGCACCCGTGACAAATGTGATAAAAAAGAAACCCCATTTTCTGAGGGGAAA 540
Db 482 TCAGAAAATTTGCACCCGTGACAAATGTGATAAAAAAGAAACCCCATTTTCTGAGGGGAAA 541
QY 541 TTCAGCTGGCTGCAGAAATTTGCATATGTAAATGAGGAGCTGAATGTTAATCCTCAAGAC 600
Db 542 TTCAGCTGGCTGCAGAAATTTGCATATGTAAATGAGGAGCTGAATGTTAATCCTCAAGAC 601
QY 601 AATGGGAAAAATATCCTCGGACATGTACAGAGTCTTCACACGACTCCATCAATCACTG 660
Db 602 AATGGGAAAAATATCCTCGGACATGTACAGAGTCTTCACACGACTCCATCAATCACTG 661
QY 661 GCCTGAGGCCCTAGGAGAAAATGGTTTTGTGGCAGAGGCCAGGGTCCCTGTGCTGTGTG 720
Db 662 GCCTGAGGCCCTAGGAGAAAATGGTTTTGTGGCAGAGGCCAGGGTCCCTGTGCTGTGTG 721
QY 721 CAGCCTAGAGACTTGGTGCCCTGTGTCGCCAGTTAATCAGCTGTGCGCTTCAGAGGGTGCA 780
Db 722 CAGCCTAGAGACTTGGTGCCCTGTGTCGCCAGTTAATCAGCTGTGCGCTTCAGAGGGTGCA 781
QY 781 AGCCCCAAGCCTTGCAGCTTCCAGTGTGTGAGCTGTGTGAGCTGTGGTGCRAAGAAGTCAAGA 840
Db 782 AGCCCCAAGCCTTGCAGCTTCCAGTGTGTGAGCTGTGTGAGCTGTGGTGCRAAGAAGTCAAGA 841
QY 841 ATTGAGGTTTGGGAACCTCCAAATCAGATTTCAAGAGATATATGGAACCCCTGGATGCC 900
Db 842 ATTGAGGTTTGGGAACCTCCAAATCAGATTTCAAGAGATATATGGAACCCCTGGATGCC 901
QY 901 AGGCAGAGTTTGTGTAGGGTGGGTGCTTCATGGAGAACCTCTGCAAGGGTAGTACAA 960
Db 902 AGGCAGAGTTTGTGTAGGGTGGGTGCTTCATGGAGAACCTCTGCAAGGGTAGTACAA 961
QY 961 AAGGGAAATGTTGGTGGGAGGCCCCACACAGAGTCCCGAGTGGGCTCCCATCTAGTAGA 1020
Db 962 AAGGGAAATGTTGGTGGGAGGCCCCACACAGAGTCCCGAGTGGGCTCCCATCTAGTAGA 1021
QY 1021 GCTGTGAGAAGAGTCCACCAATCTCCAGACTCCAGAGAGGGTAGATCCACGTACAGACTTG 1080
Db 1022 GCTGTGAGAAGAGTCCACCAATCTCCAGACTCCAGAGAGGGTAGATCCACTGACAGCTTG 1081
QY 1081 CAGCATGTGCTGAAAATCCACACACTCAGTGCAGCCCTGTGAAAGCAGCAGGGATG 1140
Db 1082 CAGCATGTGCTGAAAATCCACACACTCAGTGCAGCCCTGTGAAAAGCAGCAGGGATG 1141
QY 1141 GAGTCTGTACCCCTACAAAACCTAGTGCAGAGCTGACCAGACCGTGGGATCTACCTC 1200
Db 1142 GAGTCTGTACCCCTACAAAACCTAGTGCAGAGCTGACCAGACCGTGGGATCTACCTC 1201
QY 1201 TTGCATTGTGACCTGGACGTGAGACATGGAGTCAAAAGAGATCATTTTGGAGCTTTA 1260
Db 1202 TTGCATTGTGACCTGGACGTGAGACATGGAGTCAAAAGAGATCATTTTGGAGCTTTA 1261
QY 1261 AGATTGACTGCCCCACTGGATTTCGGACTTATATATGGGGCCCGTACCCCTTTGTTTGGC 1320
|||||

Db 1262 AGATTGACTGCCCCACTGGATTTGCGACTTATATGGGCCCGTACCCCTTTGTTTGGC 1321
QY 1321 CAATTTTTTCCATTTGGAACCTGCGTATTATTACCAATGCTGTACCTCATCTATG 1377
Db 1322 CAATTTTTTCCATTTGGAACCTGCGTATTATTACCAATGCTGTACCTCATCTATG 1378
RESULT 2
US-09-997-610-5
; Sequence 5, Application US/09997610
; Patent No. US2002015624A1
; GENERAL INFORMATION:
; APPLICANT: Fox, Brian
; APPLICANT: Holloway, James L.
; TITLE OF INVENTION: ADIPOCYTE COMPLIMENT RELATED PROTEIN
; TITLE OF INVENTION: ZACRP13
; FILE REFERENCE: 00-96
; CURRENT APPLICATION NUMBER: US/09/997,610
; CURRENT FILING DATE: 2001-11-29
; PRIOR APPLICATION NUMBER: US 60/253,924
; PRIOR FILING DATE: 2000-11-29
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 5
; LENGTH: 1731
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1731)
US-09-997-610-5
Query Match 100.0%; Score 1377; DB 9; Length 1731;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1377; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ATATGGGTGCATACCTGTCTTAATAACGGCAGTCATTGAGCATGTAGAAGTTGCTGACCT 60
Db 355 ATATGGGTGCATACCTGTCTTAATAACGGCAGTCATTGAGCATGTAGAAGTTGCTGACCT 414
QY 61 CCAGCACACCCAGGCCCCCGCCAGGAAGTGGGGCTCTCTGGTGCCAGCAGTTTACCACAA 120
Db 415 CCAGCACACCCAGGCCCCCGCCAGGAAGTGGGGCTCTCTGGTGCCAGCAGTTTACCACAA 474
QY 121 TATACAGAGAAATAGTGAAATGACAAATGCCCTGTCTCGATATACAAAGGTCAGC 180
Db 475 TATACAGAGAAATAGTGAAATGACAAATGCCCTGTCTCGATATACAAAGGTCAGC 534
QY 181 TTTACTGTGAAGCTCAGTGGAAAACCTTCCTCTTCCCTTTCAAGCCCATCATCTTCACAGG 240
Db 535 TTTACTGTGAAGCTCAGTGGAAAACCTTCCTCTTCCCTTTCAAGCCCATCATCTTCACAGG 594
QY 241 GTCCTGTACAATGCCAGAGGATTTAAAGAGGCCATGGGAGTCTTTGCTTGCAGGGTG 300
Db 595 GTCCTGTACAATGCCAGAGGATTTAAAGAGGCCATGGGAGTCTTTGCTTGCAGGGTG 654
QY 301 CCTGGGAATTTACTACTCCAGCTTTGATGTGTAGCTGCATCATTTGCAAGGTGAATATTTGG 360
Db 655 CTTGGGAATTTACTACTCCAGCTTTGATGTGTAGCTGCATCATTTGCAAGGTGAATATTTGG 714
QY 361 CTAATGAGGAACCAATTTTGCTTAATAGGAAGAAATTTCTAAGCAGCAAAAGCATTTCAA 420
Db 715 CTAATGAGGAACCAATTTTGCTTAATAGGAAGAAATTTCTAAGCAGCAAAAGCATTTCAA 774
QY 421 GAGGTGACTTGGTGTGTTAAAGGCATTCAGTTTCATAAGGAGCGCAGAGCATAGAGT 480
Db 775 GAGGTGACTTGGTGTGTTAAAGGCATTCAGTTTCATAAGGAGCGCAGAGCATAGAGT 834
QY 481 TCAGAAAATTTGCACCCGTGACAAATGTGATAAAAAAGAAACCCCATTTTCTGAGGGGAAA 540
Db 835 TCAGAAAATTTGCACCCGTGACAAATGTGATAAAAAAGAAACCCCATTTTCTGAGGGGAAA 894
QY 541 TTTCAAGCTGGCTGCAGAAAATTTGCATATGTAAATGAGGAGCTGAATGTTAATCCTCAAGAC 600
|||||

Db 895 TTCAGCTGGCTGCAGAAATTTGCATATGTAATGAGGAGCTGAATGTTAAATCCTCAAGAC 954
Qy 601 AATGGGAAAATATCTCTCGACATGTGAGAGCTTTCACAGCAGTCCATCAAAATCACTG 660
Db 955 AATGGGAAAATATCTCTCGACATGTGAGAGCTTTCACAGCAGTCCATCAAAATCACTG 1014
Qy 661 GCCTGGAGCCCTAGAGAAAATGTTTGTGGGACAGCCCGAGGTCCTGTGCTGTG 720
Db 1015 GCCTGGAGCCCTAGAGAAAATGTTTGTGGGACAGCCCGAGGTCCTGTGCTGTG 1074
Qy 721 CAGCCTAGAGACTTGGTGGCCCTGTCCACGTATTAATCAGCTGTGGCTTCAGAGGGTCA 780
Db 1075 CAGCCTAGAGACTTGGTGGCCCTGTCCACGTATTAATCAGCTGTGGCTTCAGAGGGTCA 1134
Qy 781 AGCCCCAAGCCTTGGCAGCTTCCAAAGTGGTGTGAGCCTGTGGGTGCAAGAAGTCAAGA 840
Db 1135 AGCCCCAAGCCTTGGCAGCTTCCAAAGTGGTGTGAGCCTGTGGGTGCAAGAAGTCAAGA 1194
Qy 841 ATTGAGGTTTGGGAACCTCCAAATCAGATTTTCAGAGATATATGAAACCCCTGGATGCC 900
Db 1195 ATTGAGGTTTGGGAACCTCCAAATCAGATTTTCAGAGATATATGAAACCCCTGGATGCC 1254
Qy 901 AGGCAGAGTTTGTGTAGGGTGGGTCTCTCATGGAGAACTCTGCAAGGGTAGTACAA 960
Db 1255 AGGCAGAGTTTGTGTAGGGTGGGTCTCTCATGGAGAACTCTGCAAGGGTAGTACAA 1314
Qy 961 AAGGGAATTTGGGTGGGAGCCCCACACAGAGTCCCGAGTGGGCTCCATCTAGTAGA 1020
Db 1315 AAGGGAATTTGGGTGGGAGCCCCACACAGAGTCCCGAGTGGGCTCCATCTAGTAGA 1374
Qy 1021 GCTGTGAGAAGTTCACCAATCTCCAGATCCAGAGGAGTACACATGACAGCTTG 1080
Db 1375 GCTGTGAGAAGTTCACCAATCTCCAGATCCAGAGGAGTACACATGACAGCTTG 1434
Qy 1081 CAGCATGTCCCTGAAAATCCACAGACACTCAGTCCAGCCCTGTGAAGAGCAGCGGATG 1140
Db 1435 CAGCATGTCCCTGAAAATCCACAGACACTCAGTCCAGCCCTGTGAAGAGCAGCGGATG 1494
Qy 1141 GAGTCTGTACCTCAAAACCGTAGTGGCAGAGCTGACCAAGACCGTGGGAATCTACCTC 1200
Db 1495 GAGTCTGTACCTCAAAACCGTAGTGGCAGAGCTGACCAAGACCGTGGGAATCTACCTC 1554
Qy 1201 TTGCATTCATGACCTGGAGCTGAGACATGGAGTCAAAAGAGATCAATTTGGAGCTTTA 1260
Db 1555 TTGCATTCATGACCTGGAGCTGAGACATGGAGTCAAAAGAGATCAATTTGGAGCTTTA 1614
Qy 1261 AGATTGACTCCCCACTGGATTTCGGACTTATATGGGGCCCGTACCCCTTTGTTGGC 1320
Db 1615 AGATTGACTCCCCACTGGATTTCGGACTTATATGGGGCCCGTACCCCTTTGTTGGC 1674
Qy 1321 CAATTTTTCATTTGGAACCTGCCGTATTTACCCAATGCCGTGACCTCCATTTGTATG 1377
Db 1675 CAATTTTTCATTTGGAACCTGCCGTATTTACCCAATGCCGTGACCTCCATTTGTATG 1731

RESULT 3

US-09-997-610-3
; Sequence 3, Application US/09997610
; Patent No. US200201562441
; GENERAL INFORMATION:
; APPLICANT: Fox, Brian
; APPLICANT: Holloway, James L.
; TITLE OF INVENTION: ADIPOCYTE COMPLIMENT RELATED PROTEIN
; FILE REFERENCE: 00-96
; CURRENT APPLICATION NUMBER: US/09/997,610
; CURRENT FILING DATE: 2001-11-29
; PRIOR APPLICATION NUMBER: US 60/253,924
; PRIOR FILING DATE: 2000-11-29
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: fastseq for Windows Version 3.0
; SEQ ID NO 3

LENGTH: 1377
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Degenerate polynucleotide encoding a polypeptide
; OTHER INFORMATION: of SEQ ID NO:2
; NAME/KEY: variation
; LOCATION: (1)...(1377)
; OTHER INFORMATION: Each n is independently A, T, G, or C.
; NAME/KEY: misc feature
; LOCATION: (1)...(1377)
; OTHER INFORMATION: n = A,T,C or G
US-09-997-610-3

Query Match 71.8%, Score 988.4, DB 9, Length 1377;
Best Local Similarity 58.8%, Pred. No. 3.4e-288;
Matches 810, Conservative 312, Mismatches 255, Indels 0, Gaps 0;
Qy 1 ATAGTGGTCATACCTGCTCTTAATAACGGCAGTCATTGAGCATGTAGAGTTCTCGACCT 60
Db 1 ATHGTGNTNATHCCNGTNTNATHACNGCNGTNTNATHGARCAYGTNGARCTNGCNGNCCN 60
Qy 61 CCAGCACACCCAGCCGCCAGAACAGTGGGGCTCTCTGGTGACACAGGTTTACCACAA 120
Db 61 CCNGCNCAYCCNMGCCNCCNGARGGTNGCNCNGCNGCNGCNGTNTNCNCAR 120
Qy 121 TATACAGGAGAAATAGTGAATGAAATGCCCTCTCTCTGATATAGAAAGGTCAGCC 180
Db 121 TAYACNGGARGATHWSNGARATGACNAARTGYCCNTGYCCNGAYATHGARMGNSGCN 180
Qy 181 TTTACTGTGAAGCTCAGTGGAAACTTCTCTCTTCTTCAAGCCCATCATCTTCACAGG 240
Db 181 TTYACNGTNAARYTNWSNGNAARYTNCNTTNCNTTYAARCCNATHATHTTYACNGN 240
Qy 241 GTCCTGTACATGCCAGAGGATTTAAAGGAGGCATCGGGAGTCTTCTCTGTCAGGCTG 300
Db 241 GTNYTNTAAYGCMARMGNGAYYTNAARGAGCATGGGNGTNTTYGNTGYMGNGTN 300
Qy 301 CCTGGAAATTACTCCTCCAGCTTTGATGTGAGCTGCATCATTTGCAAGTGAAATTTGG 360
Db 301 CCNGNAAYTAYTAVWSNWSNTTYGAYTNGARYTNCAYCATGYAARCTNAAATHTGG 360
Qy 361 CTAATGAGGAAGCAAAATTTTGGCTAATAGGAAGAAATTTCTAAGCAGCAAGCAATCAA 420
Db 361 YTNATGMAARCARATHYTNGCNAAYAARGARGARATHWSNAARCARCARNSATHCAR 420
Qy 421 GAGTGACTCTGGTCTCTTAAAGCATTCAGTCTTCAAGGAGCGAGCAGATGAAGT 480
Db 421 GARGTNACTGGTNTTYTNAARCNTTYNSNTTYATHMGNGARGCARGAYAAARWSN 480
Qy 481 TCAGAAAATTTGCCCTCGACAATGTATAAAAAAGAAACCCATTTTCTGAGGGGAA 540
Db 481 WSGARAAYTNCAYCCNGAYAACTNATHAARAARAARAYCCNTTYWSNARGSNAAR 540
Qy 541 TTCAAGCTGGCTGCAGAAATTTGCATATGTATAGGAGCTGAATGTTAATCCTCAAGAC 600
Db 541 TTYAARYTNGCNGCARATHGTGATHTGYAAYGARGARYTNAAYCTNAAAYCCNCARGAY 600
Qy 601 AATGGGAAAATATCTCTCGACATGTACAGAGTCTTACACAGCAGTCCCATCAATCACTG 660
Db 601 AAYGNGARAAYATHWSNTGGACNTGYCARMGNWSNCAARCARWSNTHAARWSNTN 660
Qy 661 GCCTGGAGCCCTAGGAGAAAATGTTTGTGGCAGAGCCCGAGGTCCTGTGCTGTG 720
Db 661 GCNTGCMGCCNMGNMNAARTGGTGYGNGACNGCNGCNGNWSNTNTGYTYGTYGN 720
Qy 721 CAGCCTAGAGACTTGGTGGCTCTCTCCAGTTAATTCAGCTGTGGCTTCAGAGGGTCA 780
Db 721 CARCCNMGNGAYTYTNGTNCNTGYCTNCNGTNAAYWSNCGNGTNGCNSNARGNGCN 780
Qy 781 AGCCCCAAGCCTTGGCAGCTTCCAAAGTGGTGTGAGCCTGTGGGTGCAAGAAGTCAAGA 840
Db 781 WSNCCNAARCCNTGGCARYTNCNNWSNGNGTNGARCCNTNGGNGCNAARAARWSNMG 840


```
OY 1141 GAGTCTGTACCTACAAACCGCTAGTGGCAGAGCTGACCAAGACCGGTGGGAATCTACCTC 1200
|||||
Db 1495 GARNSNGNCCNTAARACNGTNGTNGCNGARTTNACNAACNGTNGGNATHYATYTN 1554
OY 1201 TTGATTGTTCATGACCTGGAGCTGAGACATGAGTCAAAAGAGATCATTTTGGAGCTTTA 1260
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1555 YTNCAITGYCAIYAYNGAYGTWNGNCAYGGNGTNAARMNGAYCAVTTTGGNGCNTN 1614
OY 1261 AGATTTGACTCCCGCCTGGATTTCGGACTTATATGGGGCCCGTACCCCTTTTGGG 1320
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1615 MGNITGYATGYCCNACNGNTTGMNACNTAYATGGGNCNGTNCNTNTGTGTYG 1674
OY 1321 CAATTTTTCATTTGGNACTGCCGTATTACCCAAATGCCCTGTACCTCCCATGTATG 1377
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1675 CARTTYYTCNTTGGNACNGCNTTAYACNCARTGYTNTAYTNCAYTGYATG 1731

RESULT 5
US-10-026-188-6/c
: Sequence 6, Application US/10026188
: Patent No. US200201645A1
: GENERAL INFORMATION:
: APPLICANT: Zhang, Yifeng
: APPLICANT: The Regents of the University of California
: TITLE OF INVENTION: Assays for Taste Receptor Cell Specific
: FILE REFERENCE: 02307E-114910US
: CURRENT APPLICATION NUMBER: US/10/026,188
: CURRENT FILING DATE: 2001-12-21
: PRIOR APPLICATION NUMBER: US 60/259,379
: PRIOR FILING DATE: 2000-12-29
: NUMBER OF SEQ ID NOS: 8
: SOFTWARE: FastSeq for Windows Version 3.0
: SEQ ID NO 6
: LENGTH: 155074
: TYPE: DNA
: ORGANISM: Homo sapiens
: FEATURE:
: OTHER INFORMATION: human genomic region containing ltrpc5
: OTHER INFORMATION: (Human Chromosome 11p15.5 PAC clone pDJ915f1
: OTHER INFORMATION: containing KvLQT1 gene)
US-10-026-188-6

Query Match 46.6%; Score 641.6; DB 9; Length 155074;
Best Local Similarity 80.7%; Pred. No. 3e-182;
Matches 826; Conservative 0; Mismatches 159; Indels 39; Gaps 5;

OY 390 GGAAGAAATTTCTAAGCAGCAAAAGCATTCAGAGGTGACTTGGGTGCTGTAAAGGCATT 449
|||||
Db 94901 GGAAGAAATTTCTAAGCTGCAAGCTTCAAGAGGTGACTTGGGTGCTGTAAAGGCATT 94842
OY 450 CAGTTTCATAGGGAGGAGCAGCATAGAGTTAGAGAAATTTGACCCCTGACATGTGAT 509
|||||
Db 94841 CAGTTTTCAGAGGGAACACAG-ATAAAGTTCAGAAATTTGACGCTCAAAATGTGAT 94783
OY 510 AAAAAAGAAACCCATTTCTAGGGGAAATTCAGCTGCCTCAGAAATTTGCATATG 569
|||||
Db 94782 AGAAATTAACCCATTTCTAGGAGAAAGTCAAGCAGCTTCAGAAATTTGCATATG 94723
OY 570 TAATGAGGAGCTGAATGTTAATCCTCAAGACAATGGGGAATATCTCCTGGACATGCA 629
|||||
Db 94722 TAACAAGAGCCAATGTTAATCACTAAGACAATGGGGAATATCTCAGGGGATGCA 94663
OY 630 GAGCTTTCACAGCAGTCCATCAATCACTGGCCCTGGAGGCTTAGAG-AAAATGGTT 686
|||||
Db 94662 GAGACCTTAGGAGGAGCCCTCCCATCCAGGCCAGAGGCTTAGAGGCAAAATGTT 94603
OY 687 TTGTGGGAGAGCCCGAGGCTCCCTGTGCTGTACGCTAGACACTTGGTCCCTGTGT 746
|||||
Db 94602 TCCTGGAGACTGGGCCAGGGCTCCATGCTGTGTGACGCTAGGAGCTTGGTGGCCCTGTGT 94543
OY 747 CCCAG-----TTAATTCAGCTGTGGCTTCAGA 773
```

```
Db 94542 CCCAGCCGCTCTAGCCATAGCTAAAAAGGTGTCAAGGTACAGCTTGGGCTGTGGCTTCAGA 94483
|||||
OY 774 GGGTGAAGCCCGCCAGCTTGGCAGCTTCCAAAGTGGTGTGAGCCCTGTGGGTGCAAGAA 833
|||||
Db 94482 GGGTGAAGCCCGCCAACTCTTGGTAGCTTCCATGTGGTGTGAGCCCTGTGGGTGCAAGAA 94423
OY 834 GTCAGAAATTGAGGTTTGGGAACCTCCAAATCAGATTTTCAGAGAATATATGAAACCCCTG 893
|||||
Db 94422 GTCAGAAATTGAGGTTATGAAAGCTCTGCCATAGATTTTCAG-AGATGTATGAAATGCTGTG 94364
OY 894 GATGCCAGGAGCAGAAAGTTTGTCTAGGGGTGGGTTCCTCATCGGAGAACTCTGCAAGGTT 953
|||||
Db 94363 GATGTCCAGGAGCAGAAAGTTTGTCTGAGGGGTGGGTTCCTCATCGGAGAACTCTGCTAGGGC 94304
OY 954 AGTACAAAAGGAAATGTTGGGTGGGAGCCCGCCACACAGAGTCCCGACGTGGGGTCCCATC 1013
|||||
Db 94303 AGTGAGGAGGAGAAAGTGGGGTGGAGCCCGCCCATGCAAAAGTCCCGACCTGGGGCACTGCC 94244
OY 1014 TAGTAGAGCTGTGAGAAGAGTCCACCATCTCCAGACTCCAGAGAGGGTAGATCCACTGA 1073
|||||
Db 94243 TAGTGGAGCTGTGAGAAGAGGAGGCCACTGTCTCTCCAGATCCCGAATGATAGATCCATCAA 94184
OY 1074 CAGCTTGCAGCATGTGCTGAAATTCACAGACACTCAGTGCAGCCCTGTGAAAGCAGC 1133
|||||
Db 94183 CAGCTTGCAGCATGTGCTGAAAGAGTGCAGACACTCAATGCCAGCCCTGTGAAAGCAGC 94124
OY 1134 AGGATGAGTCTGTACCTACAAACCGTAGTGGCAGAGCTGACCAAGACCGTGGGAAT 1193
|||||
Db 94123 CAGGAGGGAGGCTGTACCTGCAAGAGCAGAGGGGAGAGTGTCAAGACTATGGGAAC 94064
OY 1194 CTACCTCTTGCAATGTCTATGACCTGGAGCTGAGACATGGAGTCAAAAGAGATCATTTTGG 1253
|||||
Db 94063 CCACCTCTTGATCAGCATGACCTAGATGTGAGACATGGAGTCAATGAGATCATTTTGG 94004
OY 1254 AGCTTTAAGATTTGACTGCCCCACTGGATTTGCGACTTATATGGGGCCCGTA-CCCTTTT 1312
|||||
Db 94003 AGCATTAAGATTTAACTGCCCACTGGATTTTGAACCTTGCATGGGGCTTTAGCCCTTC 93944
OY 1313 GTTTTGGCAATTTTTCATTTGGAAGTGCCTGATTTTACCAATGCTGTACCTCCATT 1372
|||||
Db 93943 GTTTTGGCAATTTACTCCCATTTTCAAAATGGGTGATTTATCCAGGCTGTACCCCTAAT 93884
OY 1373 GTAT 1376
|||||
Db 93883 GTGT 93880

RESULT 6
US-09-933-267A-1
: Sequence 1, Application US/09933267A
: Patent No. US20020123095A1
: GENERAL INFORMATION:
: APPLICANT: Kalush, Francis et al.
: TITLE OF INVENTION: Estrogen receptor alpha variants and
: FILE REFERENCE: C1000258C14
: CURRENT APPLICATION NUMBER: US/09/933,267A
: CURRENT FILING DATE: 2001-08-21
: PRIOR APPLICATION NUMBER: 60/160626
: PRIOR FILING DATE: 1999-10-20
: PRIOR APPLICATION NUMBER: 60/183756
: PRIOR FILING DATE: 2000-02-22
: PRIOR APPLICATION NUMBER: 09/692414
: PRIOR FILING DATE: 2000-10-20
: PRIOR APPLICATION NUMBER: 09/768184
: PRIOR FILING DATE: 2001-01-24
: PRIOR APPLICATION NUMBER: 09/804076
: PRIOR FILING DATE: 2001-03-13
: PRIOR APPLICATION NUMBER: 09/826314
: PRIOR FILING DATE: 2001-04-05
: NUMBER OF SEQ ID NOS: 2
: SOFTWARE: FastSeq for Windows Version 4.0
```



```
; SEQ ID NO 1
; LENGTH: 465237
; TYPE: DNA
; ORGANISM: human
US-09-933-267A-1

Query Match      44.9%; Score 618.2; DB 10; Length 465237;
Best Local Similarity 78.2%; Pred. No. 6.6e-175;
Matches 847; Conservative 0; Mismatches 193; Indels 43; Gaps 7;

QY 334 CTGCATCATTCGAAGTGAATATTTGGCTAATGAGGAACCAATTTTGGCTAATTAAGGNA 393
DB 438730 CTGGCTGTAGAGATTTGTGGAATTTGAACCTTGAGAGAAATGATTTGGGGTACCTGGTAAAA 438789

QY 394 GAAATTTTCAAGCAGCAAGCAATCAAGAGTGAAGTGGTGGTCTGTAAAGGCATTCAGT 453
DB 438790 GAAATTTTCAAGCAGCAAGCAATCAAGAGTGAAGTGGTGGTCTGTAAAGGCATTCAGT 438849

QY 454 TTCTAAGAGGAGGAGAGCATAAAGAGTTTCAGAAAAATTTGCACCCCTGACAAATGTGATAAA 513
DB 438850 TT--TAAAGAGAGAAACAGCATAAAGTTTCAGAAAAATTTGCAGCCCTGATGATCAGTAGGA 438907

QY 514 AGAAAAACCCCA-TTTTCTGAGGGGAAATTCAGCTGGCTGCAGAAAATTTGCATATGTAA 572
DB 438908 AAAAAAACCCCAATTTTGTAGAGGAAATTCAGCTGGCTGCAGAAAATTTGCATATGTAA 438967

QY 573 TGAGGAGCTGAATGTTAATCTCAAGACAAATGSGGAAAAATATCTCTCGACATGTCAGAG 632
DB 438968 CAAGGAGCCAAATGTTAATCCCAAGACAAATGGGAAAAATGTTCTCCAGAGCATGTCATAG 439027

QY 633 GTCTTCACAGCAGTCCATCAAACTACTGGCCCTGAGGCCCTAGGAG-----AAAATGGTTTT 688
DB 439028 GTCTTCATGGCAGCCCTCCCATCACAGACCGCGGAAGCCTAGGAGGAAAAAACAGTTTT 439087

QY 689 GTGGACAGGCCAGGTCCTCTGTGTGTCAGCCTAGAGACTTGTGGTCCCTGTGTGCC 748
DB 439088 GTGGCCAGTCCAGGGTCCCATGCTGTGTGACGCTAGGAACCTGTGGTCCCTGCATCT 439147

QY 749 CAGTTAAATTCAGCTGT-----GCTTCAGAGGCTGCAAGCC 784
DB 439148 CAGCTGCTCCAGCTATGTTAAAGGGGCTGAGGTACCAACGGTTTCAGAGGTTGCAAGCC 439207

QY 785 CCAAGCCTTGGCAGCTTCCAAGTGGTGTGAGCCCTGTGGGTGCAAGAGTCAAGAAATTTG 844
DB 439208 CCAAACTTGGCAGCTTTCATCTGTGTGAGCCCTGTGTGTACACAGAAATTTAAGAAATTTG 439267

QY 845 AGGTTTGGGAACCTCCAATCAGATTTTCAGAAATATATATGGAACCCCTGGATGCCAGGC 904
DB 439268 AGGTTTGGGAACCTCCAATCAGATTTATTTTCAAGATACGTGGAAATGCCCTGGATACCCAGGC 439327

QY 905 AGAAGTTTCTGTAGGGTGGGTGCTCATGAGAAACCTCTCAAGGGGTAGTACAAAAGG 964
DB 439328 AAACATTTGCTGCAGAGTGGGGCCCTCATGAGAGGCCCTCTGCTAGGGCAATAGGAAGG 439387

QY 965 GAAATGTTGGGTGGAGCCCCACACAGAGTCCCAGTGGGCTCCATCTAGTAGAGCTG 1024
DB 439388 GAAATGTTGGGTGGAAACCCCAACAGAGTCCCACCTGGGCACTGCCCTAGTGGAGCTG 439447

QY 1025 TGAGAAAGAGTCCACCATCTCCAGACTCCAGAGGGTAGATCCACTGACAGCTTGCAGC 1084
DB 439448 TAAGAGGAGGACCACTGTCTCCAGACCCGAGAATAGATATCCACTGACAGCTTGCACC 439507

QY 1085 ATGTGCCTGAAAANTCCACAGACACTCAGTGCACGCTGTGAAGCA-----GCA 1134
DB 439508 ATGTGCCTGGAANAAGCCACAGACACTCAACGCCAGCCCTGTGAAGCGAGTCAAGGTTGGAG 439567

QY 1135 GGGATGGAGTCTGTACCTTACAAAAACCGTAGTGGCAGAGCTGACCAAGACCCGTGGGAATC 1194
DB 439568 GTGGTGGTGGCTATACCTCTATAAGCCACAGGGGACAGAGCTGCCCAAGACTATGGGAACC 439627

QY 1195 TACCTCTTGCATGTGCATGACCTGGACGCTGAGACATGAGAGTCAAAAGAGATCATTTTGA 1254
DB 439628 TACCTCTTGCATCAGCATGACCTGGATGTGTGAGACATTTCAAGAGGAGAT-ATTTTGA 439686
```

```
QY 1255 GCTTTAAGATTTGACTGCCCACTGGATTCGGACTTATATGGGCCCT-ACCCCTTTG 1313
DB 439687 GCTTTAAGATTTGACTGCCCTGGTGGATTTAGACTTGTGTGGCCCTGTAAACCCCTTTG 439746

QY 1314 TTTTGGCCAAATTTTTCATTTGGAACCTGCGTATTTACCCAAATGCTGTACCTCCATTTG 1373
DB 439747 TTTTGGCCAAATTTCTCCCACTTTGGAGCTGCTGTATTTACCCAATGCTTAACCCGCAATTTG 439806

QY 1374 TAT 1376
DB 439807 TAT 439809

RESULT 7
US-10-003-806-6
; Sequence 6, Application US/10003806
; Patent No. US20020119929A1
; GENERAL INFORMATION:
; APPLICANT: Bishop, Colin E.
; APPLICANT: Agoulnik, Alexander I.
; APPLICANT: Zhu, Qichao
; TITLE OF INVENTION: CAN 1 AND ITS ROLE IN MAMMALIAN INFERTILITY
; FILE REFERENCE: P020660S1/10024824
; CURRENT APPLICATION NUMBER: US/10/003,806
; CURRENT FILING DATE: 2001-11-02
; PRIOR APPLICATION NUMBER: 60/245,872
; PRIOR FILING DATE: 2000-11-03
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 6
; LENGTH: 180557
; TYPE: DNA
; ORGANISM: Human
US-10-003-806-6

Query Match      44.3%; Score 610.2; DB 12; Length 180557;
Best Local Similarity 78.9%; Pred. No. 9.8e-173;
Matches 809; Conservative 0; Mismatches 168; Indels 48; Gaps 5;

QY 390 GGAAGAAATTTCTAAGCAGCAAGCAATTCAGAGGTGACTTGGGTGCTGTTAAAGGCATTT 449
DB 122662 GGAAGAAATTTCTAAGCAGCAAGCAATTCAGAGGTGACTTGGGTGCTGTTAAAGGCATTT 122721

QY 450 CAGTTTTCATAAGGGA---GGCAGAGCATAAAGATTCAGAAAAATTTGCACCCCTGACAAATCT 506
DB 122722 TAGTTTAAAGGGGAACAGCATAAAATTAAGTTCAAAAAATTTGCAGCCCTGCANATGT 122781

QY 507 GATAAAAAAGAAAAACCCATTTTCTGAGGGGAAATTCAGAGCTGGCTGCAGAAAATTTGCAT 566
DB 122782 GATAAAAAAGAAAAATCCCGTTTCTGAGCAGAAATTCAGGCCAGCTGCAGAAAATTTGCAT 122841

QY 567 ATGTAATGAGGAGCTGAATGTTAATCCTCAAGACAAATGGGAAAAATATCTCTCGNACATG 626
DB 122842 AAGTAACAAAGAGCCAAATGTTAGTCCCAAGATAATGGGAAAAATGTTCTCCAGGGCATG 122901

QY 627 TCAGAGGTCTTCACAGCAGTCCATCAAAATCAGTGGCTCGAGCCCTAGGAGAAAAATGTTT 686
DB 122902 CCAGAGGTCTTCACGCGACCCCACTCACATAACAGCCCTCGAGGTGTAGGAGGAAAAATG 122960

QY 687 TTGTGGGACAGGCCCGCAGGTCCTCTGTGTGTGACGCCCTAGAGACTTGGTCCCTGTGT 746
DB 122961 -----GGGCACAGGGTCCCAAGCAGTGTGTGCAGGCTAGGAGCTTGGTCCCTGCAT 123011

QY 747 CCCAGTTAATTCAGCT-----GTGCTTCAGA 773
DB 123012 CCCAGTGTCTCCAGCTGTGGCTGAAAAAGGCCAGTGTAGAGCTTGGGCCGTGGCTGCAGA 123071

QY 774 GGGTGAAGCCCAAGCCCTTGGCAGCTTCCAAAGTGGTGTGAGCCCTGTGGGTGCAAGAA 833
DB 123072 GGGTGAAGCCCAAGCCCTTGGCAGCTTCCCAAGTGGTGTGAGCCCTGTGGGTGCAAGAA 123131

QY 834 GTCAAGAATTGAGTTTGGGAACCTCCCAATCAGATTTTCAGAGATATATATGGAACCCCTG 893
```


Db 123132 GTCCAGAACTGAGGTTTCAGAAACCTCCCGCTAGATTTTCAGAGATGTTAGGAAATTCCTG 123191
Qy 894 GATGCCAGGAGAGATTTGCTGAGGGTGGGTCCTCATGAGAACTCTGCAAGGGT 953
Db 123192 GATGCCAGGAGAGATTTGCTGAGGGTGGGCTCATGAGAACTCTGTTAGGAC 123251
Qy 954 AGTACAAAGGAGATTTGGTGGGAGCCGCCACACAGAGTCCCACTGGGGTCCATC 1013
Db 123252 AATGGGAAGGAGATTTGGATTTGGATTTGCCACACAGAGTCCCTACTGAGGCACCGC 123311
Qy 1014 TAGTAGAGCTGTGAGAGAAAGTCCACCACTCTCCAGACTCCAGAGGCTAGATCCACTGA 1073
Db 123312 TAGTGTAGCTGTGAGAGAGGGGCCACCATCTTCCAGACCCAGAGATGTTAGATCCACTGA 123371
Qy 1074 CAGCTTGCAGCATGCTGCAAAATCCACAGACACTCAGTGCCAGCCTGTGAAAGCAGC 1133
Db 123372 CAGCTTGCACACTGCTGCAAGAGGCCACAGGCACTCAACACAGCCCATGAACTCAGC 123431
Qy 1134 AGGATGAGGT-CTGTACCCCTACAAACCGTAGTGGCAGAGCTGACCAAGCCGTTGGAA 1192
Db 123432 CGGGTGGGGTGCCATACCTGCAAAATCACAGGGGTCTAGCTCCCAAGGTCTATGGGA 123491
Qy 1193 TCTACTCTTGCAATGCTGACCTGAGCTGGAGCTGAGACATGGAGTCAAAAGAGATCATTTG 1252
Db 123492 CTCAACTCTTGTATCAGCATGATCTGGATGTGAGACATGGAGTCAAAAGGAGATCATTTG 123551
Qy 1253 GAGCTTTAAGATTTGACGTGCCCCACTGATTTGCGAGCTTATATGGGGCCGTA-CCCTCT 1311
Db 123552 GAGCTTTAAGATTTGATTTGCCCTGCTGGATTTGGACATGGAGTGGGGCTGTAGCTCCT 123611
Qy 1312 TGTTTTGGCCAAATTTTCCATTTGGAACCTGCCGTATTTACCCAAATGCTGTACCTCCAT 1371
Db 123612 TGTTTTGGCCAAATTTCCCTCCATTTGGAATGGCTGTATTTATCGAATGCTATACCCCTGT 123671
Qy 1372 TGTAT 1376
Db 123672 TGTAT 123676

RESULT 8

US-10-003-806-9

; Sequence 9, Application US/10003806

; Patent No. US20020119929A1

; GENERAL INFORMATION:

; APPLICANT: Bishop, Colin E.

; APPLICANT: Agoulouk, Alexander I.

; APPLICANT: Zhu, Qichao

; TITLE OF INVENTION: CAN 1 AND ITS ROLE IN MAMMALIAN INFERTILITY

; FILE REFERENCE: P0206051/10024824

; CURRENT APPLICATION NUMBER: US/10/003.806

; CURRENT FILING DATE: 2001-11-02

; PRIOR APPLICATION NUMBER: 60/245,872

; PRIOR FILING DATE: 2000-11-03

; NUMBER OF SEQ ID NOS: 14

; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 9

; LENGTH: 180557

; TYPE: DNA

; ORGANISM: Human

US-10-003-806-9

Query Match 44.3%; Score 610.2; DB 12; Length 180557;
Best Local Similarity 78.9%; Pred. No. 9.8e-173;
Matches 809; Conservative 0; Mismatches 168; Indels 48; Gaps 5;

Qy 390 GGAAGAAATTTCTAAGCAGCAAGCAATTCAGAGGTGACTTGGGTGCTGTTAAAGGCAT 449
Db 122662 GGAAGAAATTTCTAAGCAGCAAGCAATTCAGAGGTGACTTGGGTGCTGTTAAAGGCAT 122721
Qy 450 CAGTTTCAATAGGGA---GGCAGACATAAGAGTTTCAGAAATTTGCACCTCGACAAATGT 506
Db 122722 TAGTTTAAAGGGAACAGCATATAAATAAAAGTTCAAAAAATTTGCAGCCTGACAAATGT 122781

RESULT 9

US-09-835-232-7/c

; Sequence 7, Application US/09835232

; Patent No. US20020098489A1

; GENERAL INFORMATION:

; APPLICANT: Leder, Philip

; APPLICANT: Leder, Benjamin

; TITLE OF INVENTION: FORMIN-2 NUCLEIC ACIDS AND POLYPEPTIDES
; FILE OF INVENTION: AND USES THEREOF
; FILE REFERENCE: 00383/052002
; CURRENT APPLICATION NUMBER: US/09/835,232
; CURRENT FILING DATE: 2001-04-12
; PRIOR APPLICATION NUMBER: US 60/196,811
; PRIOR FILING DATE: 2000-04-13
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7
; LENGTH: 170834
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(170834)
; OTHER INFORMATION: n= A,T,C, or G
US-09-835-232-7

Query Match 43.8%; Score 603.4; DB 10; Length 170834;
Best Local Similarity 79.6%; Pred. No. 1.le-170;
Matches 816; Conservative 0; Mismatches 166; Indels 43; Gaps 7;

QY 391 GAAGAAATTTCTAAGCAGCAAAAGCATTCAGAGGTGACTTGGTGTCTGTTAAAGGCATTC 450
|||||
Db 113857 GAAGAAATTTCTAAGCAGCAAAAGCATTCAGAGGTGACTTGGTGTCTGTTAAAGGCATTC 113798

QY 451 AGTTTCATAGAGGAGCGACGACATAGAGTTTCAGAAAATTTCCAGCCCTCAGCAATGTGATA 510
|||||
Db 113797 AGTTTAAAGGAATAAGACATAAAATTTTGGAAAATTTCCAGCCTCAGCAATATGATA 113738

QY 511 AAAAAAGAAACCCATTTTCTGAGGGGAAATTCAGCTGGCTGCAGAAATTTGCATATGT 570
|||||
Db 113737 GAAAGAAAAGCTCATTTTCTGAGGAGAAATTCAGCAGGCTGCAGGAATTTGCATAAGT 113678

QY 571 AATGAGGAGCTGAATGTTAATCCTCAAGACAATGGGGAATAATCTCTGGACATGTGAC 630
|||||
Db 113677 AATGAGGAGTTGAATGTTAATCACCAGACAATGGGGTAAAGCTCTCCAGGGCATGTAC 113618

QY 631 AGGTCTTCACACAGTCCATCAATCACTGCGCTGGAGCCT---AGGAGAAATGGTTT 687
|||||
Db 113617 AGACCTTGGCAGCAGACCTTCCATCAGAGCCGACAGGCTTGGGAGGATAAAAATGGTTT 113558

QY 688 TGTGGACAGGCCAGCGTCCCTGTGCTGTGCAGCCTTAGACACTTGGTGCCTGTGTC 747
|||||
Db 113557 TGTAGCAGGCCCGAGGTCCTGCTGTGTGCAGCCTTAGGACTTGTGCTCTGCATC 113498

QY 748 C-----CAGTTAATTCAGTGTGGCTTCAGAGG 775
|
Db 113497 CAGCTGCTCTAGCCATAGCTAAAAGGGTCCAAAGGTATAAATCGGGCTGTGGCTCTAGAAG 113438

QY 776 GTGCAAGCCCCAAGCCTTGGCAGCTTCCAAGTGTGTTGAGCCTGTGGTGCAAAAGT 835
|||||
Db 113437 GTGCAAGTCCAAAGTCTTGGCAGCTTCCACGTGTGTTGAGCTTCCAGGCCACAGAAG- 113379

QY 836 CAAGAATTGAGGTTTGGGAACCTCCAATCAGATTTCAGAAAGATATATGGAACCCCTGGA 895
|||||
Db 113378 -AAGAATTGAGGTTTGGGAACCTCCGCTAGATTTACAGAGGATATATGGAATGCTGGA 113320

QY 896 TGCCCAGGCAGAAAGTTTGTGTAGGGGTGGGTCTCATATGAGAACCTCTGCAAGGGTAG 955
|||||
Db 113319 TCTCTAGGAAGAAGTTTGGCAGAGGGGACGGTCTCATATAGAGAACCTCTGTAGGGCAG 113260

QY 956 TACAAAAGGAATGTGGTGGAGGCCCCACACAGAGTCCCACTGGGGCTCCATCTA 1015
|||||
Db 113259 TGCAGAGGAATAATGGGTTTGGAGCCCCCAGACAGAGTCCCTACTGGGGCAGTGCCTTA 113200

QY 1016 GTAGAGCTGTGAAGAAGTCCACCATCTCCAGACTCCAGAAAGGTAGATCCACTGACA 1075
|||||
Db 113199 GTGGAGCTGTGAAGAAGAGGCCACTGTCTCCAG-CCCAGAAATAGTAGATCCACTGATA 113141

QY 1076 GCTTGCAGCATGTGCTGAAAAATCCACAGACACTAGTGGCAGCCTGTGAAAGCAGCAG 1135
|||||

Db 113140 GCTTGCACCATGCACCTGGAAAAGCCATAGACACTCAATGCCAGCCCATGAAAGCAGCTG 113081

QY 1136 GGATGGAGTCTGTACCTACAAAAACCGTAGTGCAGAGCTGACCAAGACCGTGGGAATCT 11195
|||||
Db 113080 GAGGGAGGCTGTACCTGCAAAAGCCACAGGGCAGAGCTGCTCAGACCATGGGAA-CC 113022

QY 1196 ACCTCTTGCAATTTGTCATGACCTGGAGCTGGAGCATGGAGTCAAAAGAGATCA---TTTTG 1252
|||||
Db 113021 ACCTCTTGCAATGAGCATAACTGGATGTGAGACATGGAGTCAAAAGGAGATCATTTTTTTG 112962

QY 1253 GAGCTTTAAGATTTGACTGCCCCCACTGGATTTCCGACATTTATATGGGGCCCGTA-CCCTTT 1311
|||||
Db 112961 GAACTTTAAGATTTGACTGCCCCCACTGGATTTCCAGACTTTGCAGGGGGCCCTGTAGCCCTT 112902

QY 1312 TGTTTTGGCCAATTTTTCATTTGGAACCTGCGGTATTTTACCAATGCTGTACCTCCAT 1371
|||||
Db 112901 TGTTTTGGCCAATTTTCTCGAATTTGGAATGGCTGTATTTTACCCATACCTATACACCAT 112842

QY 1372 TGTAT 1376
|||||
Db 112841 TGTAT 112837

RESULT 10
US-09-734-674-3/c
; Sequence 3, Application US/09734674
; Patent No. US20020081648A1
; GENERAL INFORMATION:
; APPLICANT: WEI, Ming-Hui et al
; TITLE OF INVENTION: ISOLATED HUMAN TRANSPORTER PROTEINS,
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN TRANSPORTER PROTEINS,
; FILE REFERENCE: CL001018
; CURRENT APPLICATION NUMBER: US/09/734,674
; CURRENT FILING DATE: 2000-12-13
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 202001
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(202001)
; OTHER INFORMATION: n = A,T,C or G
US-09-734-674-3

Query Match 42.9%; Score 590.4; DB 10; Length 202001;
Best Local Similarity 78.4%; Pred. No. 1e-166;
Matches 803; Conservative 0; Mismatches 171; Indels 50; Gaps 6;

QY 391 GAAGAAATTTCTAAGCAGCAAAAGCATTCAGAGGTGACTTGGTGTCTGTTAAAGGCATTC 450
|||||
Db 40027 GAAGAAATTTCTAAGCAGCAAAAGCATTCAGAGGTGACTTGGTGTCTGTTAAAGCATTC 39968

QY 451 AGTTTCATAGAGGAGCGACGACATAGAGTTTCAGAAAATTTGCACCCCTGACAAATGTGATA 510
|||||
Db 39967 GGTTTTAAAGGAACAAGGATAAAAGTTTCAGAAAATTTGCAGCCTGACAGTGCATA 39908

QY 511 AAAAAAGAAAACCCATTTTCTGAGGGGAAATTCAGCTGGCTGGCAGAAAATTTTGCATATGT 570
|||||
Db 39907 GAAAAGAAAATTCATTTTCTGAGGAGAAATTCAGCCCAATTCAGAAAATTCACAAAAGT 39848

QY 571 AATGAGGAGCTGAATGTTAATCCTCAAGACAATGGGAAAATATCTCTGGACATCTCAG 630
|||||
Db 39847 AATGAGGAGCCAAATGTTAATCCCCAAGACAATGGGAAAATATGTCTCCAGGGCATGTCAA 39788

QY 631 AGGTCTTTCACAGCAGTCCATCAAAATCACCTGGCTGGAGGCCCTAG---GAGAAAATGGTTT 687
|||||
Db 39787 AGGTCTTTATGCGAGCCCTCTCCCATCACAGCCAGGCCCCAGGAAGAAAAGTGGTTT 39728

QY 688 TGTGGGACAGGCCCGAGGGTCCCTGTGCTGTGACGCTTAGAGACTTGTGCTCCTGTGC 747
|||||

Db	39727	TCTCAGTCAGGCACAGGCTACCCATCTCTGTGTACAGCCTAGGGACCTGTGTGGCCCTGTGTT	39668
QY	748	CCAGTTAAATTCA-----GCTGTGGCTTCAGAG	774
Db	39667	TCAGCTGCTCCAGCCATAGCTGAAGGGACAAATGTAGAGCTCAGCCTGTGGCTTCAGAG	39608
QY	775	GGTGCNAAGCCCCAAGCCTTGGCAGCTTCCAAAGTGGTGTGAGCCTCTGGGTCCAAAGAG	834
Db	39607	GGTGAAGCCTCAAGCCTTGGCAGCTCCCATGTGGTGTGAGCCTCAGGTGTACGGAAG	39548
QY	835	TCAAGAATTGAGTGTGGGAACCTCCAAATCAGATTTTCAGAGATATATGAGAAACCCCTGG	894
Db	39547	TCAAGAATTGGGGTTTGAAGAACCTCTGCCTAGATTTTCAGAGGATGTATGAAATCGCTGG	39488
QY	895	ATGCCCAGCAGCAAGCTTTGCTGTAGGGGTGGGGTCCCTCATGAGAACCTCTCGAAAGGCTA	954
Db	39487	ATGCCCAGCAGCAAAATTTGCTTCAGGGCGAGGGCCCTCATGAGAACCTCTGCTAGGCCA	39428
QY	955	GTACAAAAGGGAAATGTTGGGTGGGAGCCCCCACACAGAGTCCCCAGTGGGGCTCCATCT	1014
Db	39427	GTGCAGAAAGGAAATGTGGGCTTACAGCCCC--CACAGAGTCCCTACTGGGACACCACT	39370
QY	1015	AGTAGAGCTGTGAGAAGAAATCCACATCTCTCCAGACTCCAGAGGGTAGATCCACTGAC	1074
Db	39369	AGTGGAGCTGTGAGAAAAGGCCCACTCTCTCTGACTGCAGAAATGGAAGATCCACTGAC	39310
QY	1075	AGCTTGCAGCATGTGCCTTGA AAAATCCAC--AGACACTCAGTGCACGCCCTGTGAAGCAGC	1133
Db	39309	AGCTTGAATAGTGTGTCGGAAAAGCAAGACACTCAACGCCAGCCCATGAAAGCAGC	39250
QY	1134	AGGGATGAGAGTCTGTACCCTACAAAACCGTAGTGGCAGAGCTGACCAAGACCGTGGGAAT	1193
Db	39249	CAGGAGGTGGGCTATATCTCCAAAGCCACGAGGGCAGAGCTGCCCAAGACCATCAGAA-	39191
QY	1194	CTACCTCTTGCAATGTCTAAGCTTGGACGTGAGACATGGAGATCAAAGAGATCATTTTGG	1253
Db	39190	-----CCATCAGCATGACCTGTCATGTGACACATGGAGTAAAGGAGATCATTTCCG	39140
QY	1254	AGCTTTAAGATTGACTGCCCCACTGGATTTCGGACTTATATGGGGCCGCTA-CCCCTTT	1312
Db	39139	AGCTTTAAGATTGATTGCTTGGCTTTCGATTTTGGACTTGCATTTGGGCTGTAGAGCCCTTT	39080
QY	1313	GTTTTGGCCAAATTTTTCATTTTGGAACTGCCCTATTTACCCAAATGCCTGTACCTCCCAT	1372
Db	39079	GTTTTGGCCAAATATCTCCCATTTGGAATGCTGTATTTACCTAATGCCTTTACCCAGCTT	39020
QY	1373	GTAT 1376	
Db	39019	GTAT 39016	

RESULT 11

```

US-09-263-959-1/c
; Sequence 1, Application US/09263959
; Patent No. US20020150891A1
; GENERAL INFORMATION:
; APPLICANT: Hood, Leroy E.
; APPLICANT: Rowen, Lee
; APPLICANT: Koop, Ben F.
; TITLE OF INVENTION: DIAGNOSTIC AND THERAPEUTIC COMPOSITIONS AND METHODS WHICH UTILIZE
; NUMBER OF SEQUENCES: 1279
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Seed and Berry LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: US
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25

```

Db 560186 TCCTCCAGACCCCAAGATGGTAGATCCACCACAACACTTGTGCTGTGTAAGTTCGAAAGGCC 560127
QY 1102 ACAGACACTAGTGCAGCCTGTGAAAGCAGCAGGAGTGGAGTCTGTACCCCTACAAAACC 1161
Db 560126 ACATGCACCTCAATGCTCAGCCTGTGAAAGCAGCAGCGGAGGAGACTGTACTCTGCAAGGCC 560067
QY 1162 GTATGGCAGACTGACCAAGACCGTGGGAATCTACCTCTTCATCTGCAATGTCACCTGGAC 1221
Db 560066 ACAGAGGTGGAGCTGCCAAGACCAGTGGGAACCCACCCCTTCCGCCAGCATGACCTGGAT 560007
QY 1222 GTGACATGGAGTCAAAAGAGATCATTTTGGAGCTTTTAAAGATTTTGAAGTTCGCCCACTGGA 1281
Db 560006 GTCAGACCTGGAGTCAAGAGAGATCGTTTGGAGCTTCAAAATTTGGTGGCCCACTGGA 559947
QY 1282 TTTTGGACTTATATGGGCC--CGTACCCCTTTTGTGGCCAAATTTTTCATTTGGAA 1339
Db 559946 TTTTACACTTTTATGGGACTTGTAAACCCCTTTGTTTGGCCAAATTTTCCCATTTGGAA 559887
QY 1340 CTGCGGTATTTACCAATGCCCTGTACCTCCATTTGAT 1376
Db 559886 TGGCTGTATTTATCTAATCCCTGTATCCCATTTGT 559850

RESULT 12
US-09-777-921A-3/c
; Sequence 3, Application US/09777921A
; Patent No. US20020115136A1
; GENERAL INFORMATION:
; APPLICANT: MERKULOV et al.
; TITLE OF INVENTION: ISOLATED HUMAN TRANSPORTER PROTEINS,
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN TRANSPORTER PROTEINS,
; TITLE OF INVENTION: AND USES THEREOF
; FILE REFERENCE: CL001103
; CURRENT APPLICATION NUMBER: US/09/777, 921A
; CURRENT FILING DATE: 2002-02-07
; NUMBER OF SEQ ID NOS: 126
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 69327
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(69327)
; OTHER INFORMATION: n = A,T,C or G
US-09-777-921A-3

Query Match 39.6%; Score 545.2; DB 10; Length 69327;
Best Local Similarity 77.0%; Pred. No. 2.4e-153;
Matches 727; Conservative 0; Mismatches 188; Indels 29; Gaps 4;

QY 462 GGAGGCAGAGCATAGAGTTTCAGAAAATTTGCACCTGACAATGTGATATAAAGAAAAA 521
Db 29020 GGAAGCAGAGCAAAAGTTTGTCAAATTTGCACCTGTGATGTGACAGAAAAGAAAA 28961
QY 522 CCATTTCTGAGG-GGAAATTCAGCTGGCTGCAGAAAATTTGCATATGTAATGAGGAC 580
Db 28960 TCTATTTCTGAGGAAAAATTCAGCCAGCTGCAGAAATCTGCATAAGAAACAAGGACT 28901
QY 581 TGAATGTTAATCCTCAAGACAATGGGAAAAATATCTCTGGACATGTTCAGAGTCTTCAC 640
Db 28900 GAAATGTTAATCAAGACAATGGGAAAAATGCCTCCAGGCACTTCAGAGGTCTTCAG 28841
QY 641 AGCAGTCCATAAATCACTGGCTGCAGGCCTAGGAG---AAAATGTTTTTGGGACAG 697
Db 28840 GAAGGGCCCTCCCATCATAGGCCAGGAGGAGGAAAAAATGTTTCTCTGGGCCAG 28781
QY 698 GCCAGGGTCCCTGTGCTGTGTCAGCCCTAGAGACTTGGTGCCCTGTGCCAGTTA--- 754
Db 28780 GCCAGGGCTTCTGCTGTTTGTGAGTCTCAGAGCTTGGTGCCCTGTCCATCCAGCCATG 28721
QY 755 -----ATTACGTGTGCTTCAGAGGTTGCAAGCGGTGCAAGCCCAAGCCCTT 793
_ _ _ _ _

Db 28720 CTAAAGAGAGCCAAAGTACAGCTCAGGCCATGGCTTCAGAAAGTTCAGACGCCCAAGCCCT 28661
QY 794 GGCAGCTTCCAAGTGGTGTGAGCCTGTGGGTGCAAGAAAGATCAAGAAATTTGAGTTTGG 853
Db 28660 GGCAGCTTCCAAGTGGTGTGAGTCTGCGGTACACAGAAAGATAAGAAATTAAGGTTTGG 28601
QY 854 AACCTCCAAATCAGATTTTCAAGAGATATATGGAACCCCTGGATGCCAGGCAAGATTG 913
Db 28600 ATCTCTGCTAGATTTTCAAGAGATGTATGGAATATGCTTGGATGCCAGACAGAAATTG 28541
QY 914 CTGTAGGGTGGTCTTCATGGAGAACCTCTGCAAGGTAGTCAAAAGGGAATTTG 973
Db 28540 CTGCAAGGGTGGAGCCCTTGAGGAGAACTCGCTAGGCACTGCAGAAAGGAATTTG 28481
QY 974 GGTGGAGAGCCCAACACAGAGTCCCAAGTGGGCTCCATCTAGTAGAGCTGTGAGAA 1033
Db 28480 GCTTGGAGCCTTACACAGAGTCCCAAGTGGAGCACTGCTAGTGGAGCTGTGAAAGAG 28421
QY 1034 GTCCACCATCTCCAGATCCCAAGAGGTAGATCCACTGACAGCTTGACGATGTGCCCTG 1093
Db 28420 GGCACCATCTCCAGTCCCAAGATAGTAGATCCACCACAGCTTGACCATGCACTG 28361
QY 1094 AAAATCCACAGACACTCAGTCCAGCCCTGTGAAAGCAGCAGGATGAGTCTGTACCT 1153
Db 28360 GAAAGCCCAACACACTCAACGCCAGCCTGTGAAAGAGCCAGAGGGGTCTGTACCT 28301
QY 1154 ACAAAACCGTAGTGGCAGAGCTGACCAAGACCGTGGAAATCTACCTTTGCAATTTGTCATG 1213
Db 28300 GAAAGCCCAAGAGTGCAGATGCCAAGCCATGGAGCCCACTTCTTGCAATTAGCATG 28241
QY 1214 ACCTGGAGCTGAGACATGAGTCAAAAGAGATCAATTTTGGAGCTTTAAGATTGAGTGC 1273
Db 28240 AGCTGGATGTGTGATGTGAGTCAAAGAGATCAATTTTGGAGCTTTAAGATTGACTAT 28181
QY 1274 CCACCTGGATTTCCGACTTATATGGGGCCCGTA-CCCTTTTGTGGCCAAATTTTCCA 1332
Db 28180 CCACCTAGATTTCGGACTTACTTTGGGSCCTGTAGCCCTTCATTTTGGCCAAATTTTCCA 28121
QY 1333 TTTGAACTGCCGTATTTACCAATGCCCTGTACCTCCATTTGAT 1376
Db 28120 TTTGNAACAAGTATATTTAATCAATGCCCTGTACCTCATTTGAT 28077

RESULT 13
US-09-819-607-3/c
; Sequence 3, Application US/09819607
; Publication No. US20030022337A1
; GENERAL INFORMATION:
; APPLICANT: BEASLEY, Ellen et al
; TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: CL001078
; CURRENT APPLICATION NUMBER: US/09/819,607
; CURRENT FILING DATE: 2001-03-29
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 25603
; TYPE: DNA
; ORGANISM: Human
US-09-819-607-3

Query Match 38.0%; Score 523.6; DB 9; Length 25603;
Best Local Similarity 76.7%; Pred. No. 4.4e-147;
Matches 695; Conservative 0; Mismatches 174; Indels 37; Gaps 3;

QY 508 ATAAAAAGAAAAACCATTTTCTGAGGGGAATTCAGCTGCGTGCAGAAATTTGCATA 567
Db 14623 AAAAAAAGAAAAAGAGAAAAAAGAACCCCATTTTCTGAGGAGAAATTTGCATA 14564
QY 568 TGTATGAGAGCTGAATGTTTATCTCAAGACAATGGGAAAAATATCTCTGGACATGT 627
_ _ _ _ _

Db 14563 AGCAACAGGAACTGAATATTAATCAACCAACACAAATGGGAAAAATGTCTCCAGGCGATGC 14504
Qy 628 CAGAGGTCTTACAGCAGTCCATCAAAATCACTGGCCTGGAGGCCCTAGGAG---AAAATGG 684
Db 14503 CACAGACCTTTGCGCGACGCCCTACCATCACAGACTCAGAGCCCTAGGAGGAAAAAATGG 14444
Qy 685 TTTTGTGGACAGGCCAGGCTCCTGTGCTGTGTGTGACGCTCAGAGCTTGGTGCCTGTG 744
Db 14443 TTTATGGGCGAGGTCCAGGCTCCTGTGCTGTGTGACGCTTGGGACTTCTCAGCTGC 14384
Qy 745 GTCCAGTTAATTTCAG-----CTGTGGCTTCA 771
Db 14393 ATCCAGCTGTCTCAGCCATCACTGAAGGGGCCAAGTACAGCTCAGGCTGTGGCTTCA 14324
Qy 772 GAGGTGCAAGCCCAAGCCCTGGCAGCTTCCAAAGTGTGTGAGCCCTGTGGTGCAAG 831
Db 14323 GAGGTACAAGCCCAAGCTGTGCAGCTTCCATGTGTGTGAGCCGCGAGGTGCACAG 14264
Qy 832 AAGTCAAGAAATTGAGTTTGGGAACCTTCCAAATCAGATTTTCAGAGATATATATGGAACCCC 891
Db 14263 AAGTCAAGAAATTGAGTTTGGGAGCCTCTGCCTGCATTTTCAGAGGATGTATGGAACACC 14204
Qy 892 TGGATGCCAGCAGAGAACTTTGCTGTAGGGTGGGCTTCCATGAGAACCTCTGCAAGG 951
Db 14203 TGGATGTCAGGCGAAGATTTGTGCAAGGGTGGGCTTCCATGAGAACCTCTGCTAGG 14144
Qy 952 GTAGTACAAAAGGAAATGTTGGGTGGGAGCCCCACACAGAGTCCCGAGTGGGCTCCA 1011
Db 14143 GCAATGCAGAGGAAATGTGGGTGTGTGCTTATACAAAGTCCCACTGGGCACTG 14084
Qy 1012 TCTAGTAGAGCTGTGAGAAAGTCCAGCATCTCCAGACTCCAGACTCCAGAGGGTAGATCCACT 1071
Db 14083 CCTAGTGGAGCTCTGAGAAGAGGCCACCATCTCTACACCCAGAAATGATATCCACC 14024
Qy 1072 GACAGCTTGACAGTGTGCTTGAATAATCCACAGACACTCAGTGCACGCTGTGAAAGCA 1131
Db 14023 AACAGCTTGACACATGACCTTGGAAAGCCACAGACACTCAATGCCAGCCCAAGAAAGCA 13964
Qy 1132 GCAGGATGGAGTCTGTACCTTACAAACCGTAGTGGCAGAGCTGACCAAGACCGTGGGA 1191
Db 13963 GCCGGAGGAAGGTGTACTCTGCAGAGCCACAAAGGCGAGCTGCCTTAAGACCCATGGA 13904
Qy 1192 ATCTACCTCTTGCATTTGATGACCTGGACGTGAGACATGAGATGCAAAAGAGATCATTTT 1251
Db 13903 ACCCAACTCTTGCATCAGCATGACCTGGATGTGAGAGATAGATCAAAAGATGTTCAATTT 13844
Qy 1252 GGAGCTTTAAGATTTGACTGCCACCTGGATTTCCGACTTATATGGGCCCCGTA-CCCT 1310
Db 13843 GAAGCTTTAAGATTTGACAGGCCCTGCTGGATTTTCAGACTTGCATGGGGCCCTGTAGCCCT 13784
Qy 1311 TTGTTTTGGCCAAATTTTTCATTTGAACTGGCGTATTTACCAATGCCCTGTACCTCCA 1370
Db 13783 TTGTTTGGCCAAATTTCCCATTTGGNATGGCTGAATTTACCACATGCCCTGTACCCCTA 13724
Qy 1371 TTGTAT 1376
Db 13723 TTGTAT 13718

RESULT 14

us-09-864-761-2925
; Sequence 2925, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
; FILE REFERENCE: AcomIca-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23

Query Match 38.0%; Score 523; DB 10; Length 1946;
Best Local Similarity 74.4%; Pred. No. 1.5e-147;
Matches 764; Conservative 0; Mismatches 180; Indels 83; Gaps 5;

Qy 351 GAATATTTGGCTAATGAGGAAGCAAAATTTGGCTAATAGGAAGAAATTTCTTAAGCAGCA 410
Db 174 GAATTTTGAACCTTGAACAGATGATTTAGGGTATCTGGTGAAGAAATTTCTTAAGCAGCA 233
Qy 411 AAGCATTCAGAGGTGACTTGGTGGCTTTAAAGCATTCAGTTTCATAAGGGAGGCAGA 470
Db 234 AAGCATTCAGAGGTGACTTGGTGGCTTTAAAGCATTCAGTTTCATAAGGGAGGCAGA 293
Qy 471 GCATAAGAGTTTCAGAAAATTTGCACCCCTGACAATGTGATAAAAAAGAAAAACCCATTTTC 530
Db 294 GCATAAAGTTTCAGAAAATTTGCACCCCTGACAATGTGATAAAAAAGAAAAATTCATTTTC 353
Qy 531 TGAGGGAAATTCAGCTGGCTGGCTGCAAGAAATTTGCATATGTATATGAGAGCTGAATGTTAA 590

NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Annonax Sequence Listing Engine vers. 1.1

SEQ ID NO 2925

LENGTH: 1946

TYPE: DNA

ORGANISM: Homo sapiens

FEATURE:

OTHER INFORMATION: MAP TO AC006337.2
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 20
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 14
OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 15
OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 13
OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 12
OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 12
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 12
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 10
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 14
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 13

```

Db 354 TGGAGGAAATTCAGAGTCCCTGCGAAATTTGCAATAAGGAGGCCCAATGTTAA 413
QY 591 TCCTCAAGACAATGGGAAAAATATCTCTCGACATGCTCAGAGGTCTTCAACAGCCTCCAT 650
Db 414 TCCCAAGACATGAGAAAAATGCTCCAGGCATGTCAGAGGTCTTCA-GCAGCCCTT 472
QY 651 CAAATCACTGGCTGGAGCCCTAGGAGAAATGGTTTTGTGGACAGGCCCGGCTCCCT 710
Db 473 CCCATCGCAGGCCTGGATGCCCTAGGAGAAATGGTTTTTAGGCCTGGCCAGGCTCCC 532
QY 711 GTGCTGTGTGACAGCTAGAGACTTGGTGCCCTGTGTCACAGTTAATCAGCTGTGGCTTC 770
Db 533 GTGCTGTGTGACAGCTAGGATTTGGTGACTGTCATCTACCACTCCAGCCATGGCTGA 592
QY 771 AGAGGTGTCAGCCCAAGCCCTTGCAGCTTCCAAATGGTGTTGAGCCCTGTGGTGCAAA 830
Db 593 AAGGG-----ACAAATGTAGAGCTTGAGCCATGGCTTTCAGA 628
QY 831 GAAGTCAAGAAATGAGTTTGGGAACCTCCAAATCAGATTTTCAGAAGATATATGAAACCC 890
Db 629 GGGTGAAGCCCCAGCTTTGGCAGCTCCA----- 659
QY 891 CTGGATGCCAGCGAGAAGTTGCTGTAGGGGTGGGTCTCATGTGAGAAACCTCTGCAAG 950
Db 660 -----CGTGGTGGGAGCCTCATGGAGAAACCTCTGCTAG 692
QY 951 GGTAGTACAAAAGGAAATGTTGGTGGGAGCCCCACACAGAGTCCCGAGTGGGCTCC 1010
Db 693 GCGAGTCGAAGAGGAATGTGGGTGGGAGCCTCCACAGAGTTCCTACTGGGGCACT 752
QY 1011 AFTCTAGTAGAGCTGTGAGAAGAGTCCACCAATCCTCCAGACTCCAGAAGGTAGATCCAC 1070
Db 753 GCCTAGTGGAGCTTTGAGAAGAGGCGCCACCAATCCTCCAGAGCCCAAGATGGTAGTCAA 812
QY 1071 TCACAGCTTGCAGCATGTGCGCTGAAAAATCCACAGACACTAGTCCGAGCCTGTGAAAGC 1130
Db 813 CAATAGCTTGCACACTAGTGCCTTGGAAAAAGCCACAGACACTCAACCGCAGCCTGTGAAAGC 872
QY 1131 AGCAGGAGTGAAGTGTGTACCCCTCAAAAACCGTAGTGGCAGAGCTGACCAAGACCGTGG 1190
Db 873 AGCTGGGAGGTGGCTGTGTACCTGCAAGCCATAGGGTGGAGCTGCCCAAGACCATGG 932
QY 1191 AATCTACCTCTTGCAATGTCATGACCTGGACCTGAGACATGAGATCAAAAGAGATCATTT 1250
Db 933 AACCACACCTCTTGCAATTTGCATGACCTGGATGTGAGACCTGGAGTCAAAAGAGATAATTT 992
QY 1251 TGGAGCTTTAAGATTTGACTGCCCCACTGGATTTGCGACTTATATGGGGCCCGT-ACCCC 1309
Db 993 TGGAGCATTAAGATTT-ACTGCCCCCTCTGGATTTTCAGACTTGCATAGGCCCTGTAAACCCC 1051
QY 1310 TTTGTTTTGGCCAAATTTTCCATTTGGAACCTGCGCTATTTACCAATGCCGTGTACCTCC 1369
Db 1052 TTTGTTTTGGCCAAATTTCCCAATTTGGAATGGCTATATTTATCCATACCTGTATCCCC 1111
QY 1370 ATGTGAT 1376
Db 1112 ATGTGAT 1118

RESULT 15
US-10-254-869-3
; Sequence 3, Application US/10254869
; Publication No. US20030027307A1
; GENERAL INFORMATION:
; APPLICANT: YE, Jane et al
; TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES
; FILE OF INVENTION: THEREOF
; FILE REFERENCE: CL001160DIV
; CURRENT APPLICATION NUMBER: US/10/254,869
; CURRENT FILING DATE: 2002-09-26
; NUMBER OF SEQ ID NOS: 8
```

```

; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 148567
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(148567)
; OTHER INFORMATION: n = A,T,C or G
US-10-254-869-3

Query Match          36.5%; Score 502.8; DB 9; Length 148567;
Best Local Similarity 71.5%; Pred. No. 2.4e-140;
Matches 781; Conservative 0; Mismatches 232; Indels 79; Gaps 6;

QY 352 ATATTTGGCTAATGAGGAGCAAAATTTTGGCTAATAAGGAAGAAATTTCTAAGCAGCAA 411
Db 100932 AACTTTGAACTTGAGAGAGATAATTTAGGGTATCTCATGGAAGAAATTTCTAAGCAGCAA 100991
QY 412 AGCATTCAAGAGGTGACTTTGGTGCTGTTAAAGGCATTCAGTTTCAT----- 458
Db 100992 AGCATTCAAGAGGTGACGTGGTGCTCTTAAAAACATTAAGTTTATTCATTCACAAAGA 101051
QY 459 -----AAGGAGGGCAGCAGCATAAAGATTTCAAGAAAT 489
Db 101052 TATGTTTGGAAATTAGAACTCATGTTTAAAGAAAAGCAGGAATAAAAGTTTCAGAAAT 101111
QY 490 TTGCACCCTGCAAAATGTGATAAAAGAAAACCCCATTTTCTGAGGGAAATTCAGCTG 549
Db 101112 TTATAGCCTGATGATGGAATAGAAAAGAAAACCTTATTTCTGAGGAGAAATTCAAACTG 101171
QY 550 GCTGCAGAAAATTTGCATATGTAATGAGGAGCTGAATGTTAATCTCTCAACAAATGGGAA 609
Db 101172 GCTCGGAAATTTGCATCAGTAATCAGGAGCAAAATGTTAATGGCCAGACCATGGGAA 101231
QY 610 AATATCTCTTGACATGTCAGAGGTCTTCAACAGAGTCCATCAAACTACTGCGCTGGAG 669
Db 101232 AATGCTCCAGGCACTGTCAAGGT-----AGCCCCCTCTATCAAGCCCTGACT 101282
QY 670 CCTAGA-CAAAATGTTTGTGGCAGAGCCCGAGGTCCTCTGCTGTGTCAGCCCTAG 728
Db 101283 CCTGGAGAAAAATGAGTTTCAATGGGCTGGGCCCGAGGCTTGTGCTTTCGTAGTCTCA 101342
QY 729 AGACTTGTGTGCTGTGCCAGTAAATTCAGCTGTG----- 765
Db 101343 GGACTTGTGCTGCTGCATCCCAGCTGTTCTTAAAGGGCCACATACAGTTTCAGACCAT 101402
QY 766 GCTTCAGAGGGTGCAAGCCCAAGCCTTGGCAGCTTCCAAAGTGTGAGTGTGAGCCTGTGGT 825
Db 101403 GCTTCAGAGGGTGTAAAGCAGCAAGCCTTGGTGGCTTAGCATGTTGTTGGGCTGTGGAT 101462
QY 826 GCAAAAGACTCAAGAAATTGAGGTTTGGAAACCTCCAAATCAGATTTCAAGAGATATATGA 885
Db 101463 GCACAGAGTCAAGAAATTGAGGTTTGGAAACCTCTGCGCTGGATTTTCAGAGATGATGA 101522
QY 886 AACCCTGGATGCCAGGCGCAAGATTTGTGTAGGGGTGGGTCTCTCATGGAGAACCTCT 945
Db 101523 AATGCTAGATGTCCTCGACAGAGTTGTCTACATGGCAGAGCCCTCTATGGAGAACCTCT 101582
QY 946 GCAAGGTAGTACAAAAGGAAATTTGGTGGGAGGCCCCACACAGAGTCCCGAGTGGG 1005
Db 101583 GCTAGGGCAGCGTGAAGGAAATATGGGTGGGAGAACCCACACAGAGTTCGCCACTAGG 101642
QY 1006 GCTCCATCTAGTAGAGCTGTGAGAAGAACTCCACCATCTCCAGACTCCAGAGGAGTGA 1065
Db 101643 GCACCACCTTAGTGAGCTGTGAGAAGAGGTCAACCATCTTCACAGACACAGAAATGGTAG 101702
QY 1066 TCCACTGACAGTTCAGCATGTGCTGAAAAATCCACAGACACTCAGTCCAGCAGCTGTG 1125
Db 101703 TCCACACAGTTTGCACCATGTGCTGAAAAGCTGCAGACATACATGCCAGCAATG 101762
QY 1126 AAGCAGAGGGATGGATGCTGTACCCCTACAAAACCGTAGTGGCAGAGCTGACAGAGCC 1185
Db 1126 AAGCAGAGGGATGGATGCTGTACCCCTACAAAACCGTAGTGGCAGAGCTGACAGAGCC 1185
```


THIS PAGE BLANK (uspto)

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 19, 2003, 16:09:30 ; Search time 90.9035 Seconds
(without alignments)
4645.518 Million cell updates/sec

Title: us-09-997-610-1-copy_2_1378

Perfect score: 1377

Sequence: 1 atagtgtacacgtctt.....gcctgtacctccattgtatg 1377

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued_Patents_NA.*

1: /cgn2_6/ptodata/2/ina/5A-COMB.seq.*

2: /cgn2_6/ptodata/2/ina/5B-COMB.seq.*

3: /cgn2_6/ptodata/2/ina/6A-COMB.seq.*

4: /cgn2_6/ptodata/2/ina/6B-COMB.seq.*

5: /cgn2_6/ptodata/2/ina/PCTUS-COMB.seq.*

6: /cgn2_6/ptodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	439.6	31.9	49136	4	US-09-422-869-1
2	282.8	20.5	44453	4	US-09-146-053-5
c	271.2	19.7	14855	2	US-08-687-080-59
3	247.4	18.0	620	4	US-08-991-789A-29
5	247.4	18.0	620	4	US-09-062-451-29
6	247.4	18.0	620	4	US-09-598-326-29
7	240.2	17.4	152331	3	US-09-128-155-16
8	204	14.8	14855	2	US-08-687-080-59
c	57.2	4.2	168575	4	US-09-426-290-1
10	48.8	3.5	7218	1	US-08-232-463-14
c	48.4	3.5	1313	2	US-08-463-911-6
12	48.4	3.5	4517	4	US-09-140-804-9
13	42.4	3.1	1107	3	US-09-188-930-217
14	40	2.9	1276	2	US-08-463-911-1
15	38	2.8	1755	4	US-08-543-246B-8
16	37.8	2.7	1063	4	US-09-490-818-6
17	37.4	2.7	728	4	US-09-336-536-2
18	37.4	2.7	1338	4	US-09-336-536-1
19	37.4	2.7	1347	4	US-09-140-804-1
20	36.6	2.7	3228	4	US-09-605-785-701
21	36.2	2.6	729	4	US-09-140-804-10
22	36	2.6	1743	4	US-08-487-596-9
23	35.8	2.6	1074	2	US-08-627-151A-15
24	35.8	2.6	1404	6	5171840-8
25	35.8	2.6	1404	6	5480796-8
26	35.8	2.6	1486	4	US-08-795-473B-3
27	35.8	2.6	1486	4	US-09-439-856-3

ALIGNMENTS

RESULT 1

US-09-422-869-1

; Sequence 1, Application US/09422869

; Patent No. 6235481

; GENERAL INFORMATION:

; APPLICANT: POLONSKY, KENNETH S.

; APPLICANT: HORIKAWA, YUKIO

; APPLICANT: ODA, NAOHISA

; APPLICANT: COX, NANCY J.

; APPLICANT: SREENAN, SEAMUS

; APPLICANT: ZHOU, YUN-PING

; APPLICANT: OTANI, KENICHI

; APPLICANT: HANIS, CRAIG L.

; APPLICANT: BELL, GRAEME I.

; TITLE OF INVENTION: METHODS OF TREATMENT OF TYPE 2 DIABETES

; FILE REFERENCE: ARCD:307

; CURRENT APPLICATION NUMBER: US/09/422.869

; CURRENT FILING DATE: 1999-10-21

; EARLIER APPLICATION NUMBER: 60/134.175

; EARLIER FILING DATE: 1999-05-13

; NUMBER OF SEQ ID NOS: 30

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 1

; LENGTH: 49136

; TYPE: DNA

; ORGANISM: Human

US-09-422-869-1

Query Match 31.9%; Score 439.6; DB 4; Length 49136;

Best Local Similarity 72.0%; Pred. No. 1.1e-129; Indels 62; Gaps 6;

Matches 684; Conservative 0; Mismatches 204;

Qy 437 TGTAAAGGCATTCAGTTTCATAGGAGGAGAGCAGATGAAGATTCAGAAAATTTGCACC 496

Db 47377 TGGATTGGAACTCACATTTAAAGGAGGAGCAGATGAAGATTTAGAAAATTTGCATC 47436

Qy 497 CTGCAATGTGATAAAAAAAGAAACCCATTTCTGAGGGGAAATTCAGCTGGCTGCAG 556

Db 47437 CTGACTATGGATAGAAAAGAAACCCATTTCTGAGGAGAAATTCAGCCAGCTGCAG 47496

Qy 557 AAATTTGCATATGTAATCAGGAGCTGAATGTTAACTCTCAACACAAATGGGAAATATCT 616

Db 47497 AAATTTGCATATGTAATCAGGAGCTGAATGTTAACTCTCAACACAAATGGGAAATATCT 47556

Qy 617 CCTGGACATGTACAGAGGCTTTCACAGCAGCTCCATCAATCACTGGCGCTGGAGCCCTAGGA 676

Db 47557 CCAGGGCATGTACAGAGGCTTTCACAGCAGCTCCATCAATCACTGGCGCTGGAGCCCTAGGA 47616

Qy 677 G---AAAATGTTTGTGGGACAGCCCGAGGTCCTCTGTGTGTGACGCTAGAGACT 733

Db 677 G---AAAATGTTTGTGGGACAGCCCGAGGTCCTCTGTGTGTGACGCTAGAGACT 733

28	35.8	2.6	2061	6	5171840-1	Patent No. 5171840
29	35.8	2.6	2061	6	5480796-1	Patent No. 5480796
30	35.8	2.6	3319	4	US-08-795-473B-2	Sequence 2, Appl
31	35.8	2.6	3319	4	US-09-439-856-2	Sequence 2, Appl
32	34	2.5	459	4	US-09-415-551-2	Sequence 2, Appl
33	34	2.5	486	4	US-09-415-551-1	Sequence 1, Appl
c	34	2.5	486	4	US-09-065-474-136	Sequence 136, App
35	33.4	2.4	301	3	US-09-065-474-136	Sequence 137, App
c	33.4	2.4	301	3	US-09-557-034-136	Sequence 136, App
36	33.4	2.4	301	4	US-09-557-034-136	Sequence 137, App
37	33.4	2.4	301	4	US-09-029-348-20	Sequence 20, Appl
38	33.2	2.4	1881	4	US-09-449-218D-1	Sequence 1, Appl
39	33.2	2.4	2301	4	US-09-449-218D-3	Sequence 3, Appl
40	33.2	2.4	2301	4	US-09-449-218D-5	Sequence 5, Appl
41	33.2	2.4	2301	4	US-09-449-218D-7	Sequence 7, Appl
42	33.2	2.4	9301	4	US-09-449-218D-18	Sequence 18, Appl
43	33.2	2.4	660	1	US-08-186-529-3	Sequence 3, Appl
44	32.4	2.4	660	1	US-08-640-386A-3	Sequence 3, Appl
45	32.4	2.4	660	1	US-08-640-386A-3	Sequence 3, Appl

Db 47617 GGAAAAATGTTTTGTGCTGGGCCCCAGGGCCCTGCTGTTGTGTCAGTCTCAGGACT 47676
QY 734 TGGTGCCTGTGTCGCCAGTTAATTACAGCTGTGCCTTCAGAGGGTGCAAGCCCCAAGCCCTT 793
Db 47677 TGGTGCCTCACATCCAG-----CAGTGGCTA 47703
QY 794 GCGAGCTTCCAAGTGGTGTGAGCCTGTGGTGCAAGAAGTCAAGAAATTGAGGTTTGG 853
Db 47704 AAAGGGGCCAATGTACAGCTTAGACCTTGTCTCAGAGGTG-CAAGCCCCAAGCCCTGGT 47762
QY 854 ACCTCCCAATCAGATTTCAGAGATATATGGAACCCCTGGATGCCAGGCGAGAGTTTG 913
Db 47763 GGCCTTACATGTGGTGTGGCCCTGCAGAT-----ACACAGAAATTTG 47804
QY 914 CTGTAGGGTGGGGTCTCATGGAGAACCTCTGCAAGGTAGTACAAAGGGAATGTTG 973
Db 47805 CTGCACTGGTGAACCTCATGTAGAACCTCTGCTAGGCACTGTAGAAGTATATGTTG 47864
QY 974 GGTGGAGCCCCACACAGAGTCCCCAGTGGG-----CTCCATCTAGTAGAGCTGTGA 1027
Db 47865 GGTGGAGCCCCCACACAATCCCACTGGGGCACTGCCTACTGCTACTGGAACCTGTGA 47924
QY 1028 GAAGAAGTCCACCACTCCACAGTCCAGAGGGTAGATCCACTGACAGCTTGCAGCATG 1087
Db 47925 GAAGAAGGCCACCACTCCACAGACCCCAAGATGGTAGATCCACTGATGGCTTGAACCATG 47984
QY 1088 TGCTGAAAAATCCACAGACACTCAGTGCAGCCCTGTGAAGCAGCAGGGATGGAGTCTG 1147
Db 47985 CACCTGGAAAGCCACAGACACTCAACACGACCTGTGAAGCGACTGGAAGGAGGCTG 48044
QY 1148 TACCCTACAAAACCTAGTGGCAGAGCTGACCAAGACCGTGGGAATCTACCTCTTGCATT 1207
Db 48045 TACCCTGCAAAAACAGAGGCGAGAGTGCCTCAAGTATGGAGCCCACTCTTGCATG 48104
QY 1208 GTCATGACCTGACGTGAGACATGAGTCAAAAGAGATCATTTTGGAGCTTTAAGATTG 1267
Db 48105 AGCCCTGACTTGAATGTGACATGGAGTCAAAAGAGATCATTTTGGAGCTTTAAGTTTG 48164
QY 1268 ACTGCCCCACTGATTTCGGACTTATATGGGCCCGT-ACCCTTTGTTTGGCCCAATTT 1326
Db 48165 ACTGCCCACTGGATTTCGGAGTTCATGAGGCGCTGTGGCCCTTCATTTTGGCCCAATTT 48224
QY 1327 TTTCCATTGGAACTCCGCTATTTACCAATGCCTTACCTCCATTGAT 1376
Db 48225 ATCCCAATTGGAAATGGGTATATTTACCAATGCCTGACCCCAATTCAT 48274

RESULT 2

US-09-146-053-5
; Sequence 5, Application US/09146053A
; Patent No. 6399349
; GENERAL INFORMATION:
; APPLICANT: Ryan, James W.
; APPLICANT: Sprinkle, Terry Joe Curtis
; APPLICANT: Venema, Richard C.
; TITLE OF INVENTION: Human Aminopeptidase P Gene
; FILE REFERENCE: MCG103
; CURRENT APPLICATION NUMBER: US/09/146,053A
; CURRENT FILING DATE: 1998-09-02
; EARLIER APPLICATION NUMBER: 60/057,854
; EARLIER FILING DATE: 1997-09-02
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
; LENGTH: 44453
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-146-053-5

Query Match 20.5%; Score 282.8; DB 4; Length 44453;
Best Local Similarity 62.2%; Pred. NO. 1.8e-79;
Matches 585; Conservative 0; Mismatches 332; Indels 23; Gaps 8;

RESULT 3

US-08-687-080-59/c
; Sequence 59, Application US/08687080
; Patent No. 5965427
; GENERAL INFORMATION:
; APPLICANT: Gregory Dolganov
; TITLE OF INVENTION: Human RAD50 Gene and Methods of Use Thereof
; NUMBER OF SEQUENCES: 175

QY 444 GGCATTTCAGTTTCAATAGGGAGGCGAGCATAGAGTTTCAGAAAAATTTGCACCTGACAA 503
Db 38135 GGAATTTTATTAATGGAGCAGAGTGTAAAGTTTGGAAAAATTTGCACCTCACTA 38194
QY 504 TGTGATAAAAAAGAAAAACCCATTTTCTGAGGGAAAAATTTCAAGCTGGCT-----GCAGAA 558
Db 38195 TGTGATAGAAAAAGAAAAAGCCCAATTTTCAGGGGATGAATTCAGCAGACTGTGGGTAACC 38254
QY 559 ATTTGCATATGTAATCAGGAGCTGAATGTATCTCTCAAGACAATGGGAAAAATATCTCC 618
Db 38255 ACTGTGTAGAGATATTGTCATAACTAAAAAGACATTTTCAGAGATCTAAGAGCAGCCCC 38314
QY 619 TGGACATGTTCAGAGGTCTTTCACAGCAGTCCATCAATCAATCACTGGCCTGAGGCTTAGAGA 678
Db 38315 TCCCATCAGAGCTCTGAGGCCAAGGTGTTTCATGGTCCAGGCTTAGGCCCTGCTGCC 38374
QY 679 AAATGGTTTTTGGGACAGGCCAGGGTCCCTGTGTGTGCGAGCCTAGAGACTTTGGTG 738
Db 38375 CTGAGCAGGCTCTGGACACTGCTTCTCGCATCCTGCCACTCCAGCTCCAGCCTTGCTG 38434
QY 739 CCTGTGTCCCACTTAA--TTCAGCTGTGCTTTCAGAGGTGCAAGCCCCAAGCCTTGGC 796
Db 38435 AAAGTCCCAAGATCAGCTCAGGTTGCTCTTCAGAAAAGTGAAGCTGTACGCCCTTGGC 38494
QY 797 AGCTTCCAAAGTGTGTGAGCCTGTGGGTGCAAGAAGTCAAGA---ATTGAGTTTGGG 853
Db 38495 AGCTTCTGTGTCATTAAGCCTGTAGGTGTGCAAGTGCAGAGTGAAGAGGCTTGGC 38554
QY 854 AACCTCCAATCAGATTTCAGAGATATATGGAACCCCTGGATGCCAGGCGAGAAAGTTTG 913
Db 38555 AGCCTCTGCTAGATTTCAGAAAGTGTATAGGAA-ATCTGGGTGCCAGGCGAG-AGCCTG 38612
QY 914 CTGTAGGGTGGGTCTCATGAGAACCTCTCCAGGGTAGTACAAAGGGAATGTTG 973
Db 38613 CTACAGGGGCGAGGCCCTCATGGAAAACTCTGCTAGGACAGTGCAGAGGGGAAACGTG 38672
QY 974 GGTGGAGGCCCCACACAGAGTCCCCAGTG-GGGTCCATCTAGTAGAGCTGTGAGAAGA 1032
Db 38673 GGTGGAGGCCCCATGTACAGTCTCCACTGAGGGCACTGCCTAGTGGGCTGTGGAAGG 38732
QY 1033 AGTCACCAATCTCCAGACTCCAGAGGTAGATCCACTGACAGCTTGCAGCATGTGCCT 1092
Db 38733 GGGCCCCGCTCTCCAAACCTCAGAATGTTAGATCCCAACAGCTCGCACCTTGCACCT 38792
QY 1093 GAAAAATCCACAGACACTCAGTCCAGCCTGTGAAGCAGCAGGGATGAGTGTGTACC 1152
Db 38793 GGAAGAGCCACAGGCACTCAACAACGTG-----TGAAAGCAGCTACAGGTGCCGAACCC 38845
QY 1153 TACAAAACCGTAGTGGCAGAGCTGACCAAGACCGTGGGAATCTACCTCTTGCATTGTCAT 1212
Db 38846 AGAAAAGCTACAGGGCGCAGAGCTGCCAAGGCTTGGAGGCCACCCCTTGTCTCAGTGT 38905
QY 1213 GACCTGGAGCTGAGACATGGAGTCAAAAGAGATCATTTTGGAGCTTTAAGATTT---GAC 1269
Db 38906 GGCCTGGATGTGAGACCTGGAGTTGAAGGAGATTAATTTTGGAGCTTTGACATTTAAAGAC 38965
QY 1270 TCGCCCACTGGATTTCGGACTTATATGGGCCCGCTACCCCTTTGTTTGGCCAAATTTT 1329
Db 38966 TACCCTGTGGGTTTCAAACTTGCATGGGGCCTGTAGGCCCTTATTTTGGCCAAATTTCTC 39025
QY 1330 CCATTGGAACTGCCGTATTTTACCCCAATGCCTGTACCTCC 1369
Db 39026 CTTTGGAAATGAAATGTTTACTCAATGCCTATCCCCC 39065

RESULT 3

US-08-687-080-59/c
; Sequence 59, Application US/08687080
; Patent No. 5965427
; GENERAL INFORMATION:
; APPLICANT: Gregory Dolganov
; TITLE OF INVENTION: Human RAD50 Gene and Methods of Use Thereof
; NUMBER OF SEQUENCES: 175

```

CORRESPONDENCE ADDRESS:
ADDRESS: Dehlinger & Associates
STREET: 350 Cambridge Avenue, Suite 250
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94306
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/687,080
FILING DATE: 17-JUL-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/592,126
FILING DATE: 26-JAN-1996
ATTORNEY/AGENT INFORMATION:
NAME: Sholtz, Charles K.
REGISTRATION NUMBER: 38,615
REFERENCE/DOCKET NUMBER: 4600-0111.30
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 324-0880
TELEFAX: (415) 324-0960
INFORMATION FOR SEQ ID NO: 59:
SEQUENCE CHARACTERISTICS:
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
INDIVIDUAL ISOLATE: 5' END OF INTRON 2 OF RAD50 GENOMIC
INDIVIDUAL ISOLATE: SEQUENCE
US-08-687-080-59

Query Match          19.7%; Score 271.2; DB 2; Length 14855;
Best Local Similarity 77.3%; Pred. No. 4.9e-76;
Matches 385; Conservative 0; Mismatches 98; Indels 15; Gaps 4;

QY   890 CCTGGATGCCAGGACAGAAGTTTCTGTAGGGTGCGGTCTCATGGAGAACCTCTGCAG 949
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB   14854 CCTGGATGCCAGGACAGATGTTGCTCAGGGTGCGGTCTCATGGAGAACCTCTGCTA 14795

QY   950 GGGTAGTACAAGAGGAATAATGTTGGGTGGAGCCCCCACACAGATCCCAGTGGGGCTC 1009
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB   14794 GGGCAATATGAAGGGAATGTTGGGTTGAACCC--ACAGAGTTCCTATGGAGGGGAC 14737

QY   1010 CATCTACTAGAGCTGCAGAGAGTCCACATCCTCCAGACT-----CCAGAG 1059
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB   14736 TGCCTAGTAGAGCTGTGAGAAGACAGCCATGTCCTCCAGACTGGTAGATCCCCCAGAA 14677

QY   1060 GGTAGATCCAAGTGCAGATGTCCTGCTGAAAAATCCACAGACACTCAGTGGCAG 1119
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB   14676 AATAGATCCAAGTGCAGATGTCCTGCTGAAAAATCCACAGACACTCAGTGGCAG 14617

QY   1120 CCGTGTAAAGACAGAGGAGTGTGTACCCCTACAAAACCCGTAAGTGGCAGAGCTGACC 1179
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB   14616 CCGTGTAAAGACAGAGGAGGCTATACCCCTGCAAAGCC--AGAAGTGGAGCTGCC 14559

QY   1180 AAGACCGTGGGAATCTACCTCTTCGATGTCATGACCTGAGCATGAGACATGGAGTCAA 1239
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB   14558 AAGGCCATGGAAGCCACCTCTTCGATCAGAGTGAACCTGGATGTGAGACATGGAGTCAA 14499

QY   1240 AGAGATCATTTTGGAGCTTTAAGATTTTGAAGTGTGAGTGTGAGTGTGAGTGTGAGT 1299
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB   14498 GGAGATCATTTTGGAGCTTTAAGATATACACCTTGCCCCCTGAATTTTCGAGCTTGCC 14439

QY   1300 CCCGTA-CCCTTTGTTTTGGCCAAATTTTTTCCATTTTGGAACTGCCCTGATTTACCCCA 1358

```

Db 242 GAACACTGCTTAATGAACTGTGAGAGATGGCCACTGTCTATCCAGACACCAAGATGATA 301
QY 1064 GATCCACTGACAGCTTGCAGCAT -GTGCCTGAAATAATCCACAGACACTCAGTCCAGGCC- 1121
Db 302 GACCCACCAAAACCTATGCCATATGCTTATATAAACCTTACAGACACTCAATGCCAGCCC 361
QY 1122 --TGTGAAGACGAGGATGAGTCTGTACCTTACAAACCGTAGTGCAGAGCTGACC 1179
Db 362 CATGAAAAAAACTGAGAAGAGACTGTNCCCTACAATGCCACGGACAGAACTGCC 421
QY 1180 AAGACGCTGGGATCTACTCTTGCATTGTCTATGACCTGGAGG-TGAGACATGGAGTCAA 1238
Db 422 CAGGCGATGGAAGACAGCTCTATATCAATGTGACCTGGATGTGAGACATGGATCCN 481
QY 1239 AAGAGATCATTTTGGAGCTTTAAGATT---TGACTGCCCCACTGGATTTCGGACTTATA 1294
Db 482 ANGAATCTNTTAAACTTCCACGGTTNAATGACTGCCCTATTANATTGAACTTANA 541
QY 1295 T-GGGGCCGCTACCCCTTTGTTTGGCCAAATTTTTCATTTGGAACTGCCCTATTACC 1353
Db 542 TCNCGCCTGTGACCTCTTTGCTTTGGCCATTCCCCCTTTTGGAAATGCTNTTTTTC 601
QY 1354 CAATGCTGTACTCCAT 1371
Db 602 CCATGCTGTNCCCTCTT 619

RESULT 5
US-09-062-451-29
; Sequence 29, Application US/09062451
; Patent No. 6344550
; GENERAL INFORMATION:
; APPLICANT: Frudakis, Tony N.
; APPLICANT: Smith, John M.
; APPLICANT: Reed, Steven G.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
; TREATMENT AND DIAGNOSIS OF BREAST CANCER
; NUMBER OF SEQUENCES: 297
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/062,451
; FILING DATE: 04-APR-1997
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.419C2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 29:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 620 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-09-062-451-29

Query Match 18.0%; Score 247.4; DB 4; Length 620;
Best Local Similarity 68.6%; Pred. No. 3e-69;
Matches 424; Conservative 0; Mismatches 177; Indels 17; Gaps 6;

QY 771 AGAGGGTGAAGCCCAAGCCCTGGCAGGTTCCAACTGGTGTGAGCCGTGGTGGTCAAA 830
Db 2 AGAGGGTACCAGCCCAAGCCCTGACAACTTCCATAGGTGTCAAGCCTGTGGTGGTCA 61
QY 831 GAAGTCAAGAAATGAGGTTTGGCAACTCCAATCAATTTTCAGAAATATATATGA---- 885
Db 62 GAAGTCAAGAAATGAGGTTTGGGATCCTCAGCCTAGATTTCAGAGGATATAAAGAAAC 121
QY 886 --AACCCTGGATGCCAGGACAGAACTTTGCTGTAGGGTGGGTCCTCATGAGAACCT 943
Db 122 CTAACACCTAGATATTCACACAAAGTTTACTACAGGATGAAGCTTTCAGGAAACCT 181
QY 944 CTCGAAGGGTAGTACAAAAGGAAATTTGGTGGGAGCCGCCACACAGAGTCCCGAGT 1003
Db 182 CTACTAGGAAGTACAGAAGAAATGTGGTTTGGAGCCCAACACAGAAATCCCTCTA 241
QY 1004 GGGCTCCATCTAGTAGAGCTGTGAGAAGAGTCCACCATCTCCAGACTCCAGAGGGTA 1063
Db 242 GAACACTGCTTAATGAACTGTGAGAAGATGGCCACTGTCTCCAGACACCAAGATGATA 301
QY 1064 GATCCACTCAGCTTGCAGCAT -GTGCCTGAAATAATCCACAGACACTCAGTCCAGGCC- 1121
Db 302 GACCCACCAAAACCTTATGCCATATTTGCCATATAAAACCTTACAGACACTCAATGCCAGCCC 361
QY 1122 --TGTGAAGACGAGGATGGAGTCTGTACCTTACAAAACCGTAGTGGCAGAGCTGACC 1179
Db 362 CATGAAAAAAACTGAGAAGAGACTGTNCCCTACAATGCCACGGACAGAACTGCC 421
QY 1180 AAGACGCTGGGATCTACTCTTGCATTGTCTATGACTGGAGG-TGAGACATGGAGTCAA 1238
Db 422 CAGGCGATGGAAGACAGCTCTTATATCAATGTGACCTGGATGTGAGACATGGAAATCCN 481
QY 1239 AAGAGATCATTTTGGAGCTTTAAGATT---TGACTGCCCCACTGGATTTCGGACTTATA 1294
Db 482 ANGAATCTNTTAAACTTCCACGGTTNAATGACTGCCCTATTANATTGAACTTANA 541
QY 1295 T-GGGGCCGCTACCCCTTTGTTTGGCCAAATTTTTCATTTGGAACTGCCCTATTACC 1353
Db 542 TCNCGCCTGTGACCTCTTTGCTTTGGCCATTCCCCCTTTTGGAAATGCTNTTTTTC 601
QY 1354 CAATGCTGTACTCCAT 1371
Db 602 CCATGCTGTNCCCTCTT 619

RESULT 6
US-09-598-326-29
; Sequence 29, Application US/09598326
; Patent No. 6423496
; GENERAL INFORMATION:
; APPLICANT: Frudakis, Tony N.
; APPLICANT: Smith, John M.
; APPLICANT: Reed, Steven G.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
; TREATMENT AND DIAGNOSIS OF BREAST CANCER
; NUMBER OF SEQUENCES: 247
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Seed Intellectual Property Law Group PLLC
; STREET: 701 Fifth Avenue, Suite 6300
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/598,326
; FILING DATE: 20-Jun-2000
; CLASSIFICATION: <Unknown>
; ATTORNEY/AGENT INFORMATION:

```
; NAME: Potter, Jane E.R.
; REGISTRATION NUMBER: 33,332
; REFERENCE/DOCKET NUMBER: 210121.419D1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 29:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 620 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 29:
US-09-598-326-29

Query Match      18.0%; Score 247.4; DB 4; Length 620;
Best Local Similarity 68.6%; Pred. No. 3e-69;
Matches 424; Conservative 0; Mismatches 177; Indels 17; Gaps 6;

Qy 771 AGAGGGTGCAAGCCCAAGCCCTGGCAGCTTCCAAGTGGTGTTCAGAGCTGTGGGTGCAAA 830
Db 2 AGAGGGTACCAGCCCAAGCCCTGACAACTTCCATAGGGTGTCAAGCTGTGGGTGCACA 61

Qy 831 GAAGTCAGAATTCAGGTTTGGGAACCTCCATCAGATTTTCAGAAATATATGA----- 885
Db 62 GAAGTCAAAATTCAGTTTTGGGATCCTCAGCCTAGATTTTCAGAGGATATATAAGAAACAC 121

Qy 886 --AACCCCTGGATCCCAAGGAGAGTTCGTGTAGGGTGGGTCTCATGAGAACTT 943
Db 122 CTACACCTAGATATTCAGACAAAGTTTACTACAGGATGAAGCTTTTCAGGAAACCT 181

Qy 944 CTGCAAGGTAGTACAAAGGAAATGTGGTGGGAGCCCAACACAGAGTCCCAAGTG 1003
Db 182 CTACTAGAAAGTACAGAAGAGAAATGTGGTGGGAGCCCAACACAGAACTCCCTCTA 241

Qy 1004 GGGCTCCATCTAGTAGCTGTGAGAAGAGTCCACCATCTCTCAGACTCCAGAGGTA 1063
Db 242 GAACACTGCCTAATGAACCTGTGAGAAGTGGCCACTGTCTATCCAGACACCAAGATGATA 301

Qy 1064 GATCCACTGACAGCTTGCAGACT--GTGCTGAAAAATCCACAGACACTCAGTGCCAGCC- 1121
Db 302 GACCCACCAAAACTTATGCCATATTGCTATAAACTCAGACACTCAATGCCAGGCC 361

Qy 1122 --TGTGAAGCAGCAGGATGGAGTGTGTACCTTACAAACCCCTAGTGGCAGAGCTGACC 1179
Db 362 CATGAAAAAAACTGAGAAGAACTGTGCCCTTACAAATGCCAGCGGAGCAGAACTGCC 421

Qy 1180 AAGACCTGGGAATCTACCTCTTCATGTGATGATGACCTGGACG--TGAGACATGGAGTCAA 1238
Db 422 CAGGCCATGGAAAGCAGACAGCTCTTATCAATGTGACCTGGATGTGAGACATGGAATCCN 481

Qy 1239 AAGAGATCATTTTGGAGCTTTAAGATT---TGACTGCCCACTGGATTTTCGACTTATA 1294
Db 482 ANGAATATNTTTTAAACTTCCACGGTTNAATGACTGCCCTATTANATTCGAACTTANA 541

Qy 1295 T-GGGGCCCTTACCCCTTTGTTTGGCAATTTTTCATTTGGAAGCTGCGGTATTTACC 1353
Db 542 TCCNGCCCTGTGACCTCTTTCCTTTGGCCATTCCTCCCTTTTGGAAATGGCTNTTTTTC 601

Qy 1354 CAATGCTGTACCTCCAT 1371
Db 602 CCATGCTGTNCCCTCTT 619

RESULT 7
US-09-128-155-16
; Sequence 16, Application US/09128155
; Patent No. 6117654
; GENERAL INFORMATION:
; APPLICANT: Pan, Yang
; TITLE OF INVENTION: NOVEL MOLECULES OF TANGO-77 RELATED PROTEIN FAMILY
; TITLE OF INVENTION: AND USES THEREOF
; FILE REFERENCE: 09404/052001
```

```
; CURRENT APPLICATION NUMBER: US/09/128.155
; CURRENT FILING DATE: 1998-08-03
; EARLIER APPLICATION NUMBER: US 60/091.650
; EARLIER FILING DATE: 1998-07-02
; EARLIER APPLICATION NUMBER: US 60/054.646
; EARLIER FILING DATE: 1997-08-04
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 16
; LENGTH: 152331
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(152331)
; OTHER INFORMATION: n = A,T,C or G
US-09-128-155-16
```

```
Query Match      17.4%; Score 240.2; DB 3; Length 152331;
Best Local Similarity 59.4%; Pred. No. 1.7e-65;
Matches 572; Conservative 0; Mismatches 320; Indels 71; Gaps 7;

Qy 444 GGCATTTCAGTTTCATAGGGAGGAGCAGACATAAGAGTTTCAGAAAATTTTGCACCCCTGCACA 503
Db 71252 GGAACCTATGTTTAAATGGAAGTAGAGTCTAAAAATTTGAAAAATTTTGCAGCCCTGGCCT 71311

Qy 504 TGTGATAAAAAAGAAAACCCATTTTCTGAGGGGAAATTCAGCTGCTGCAGAAATTTG 563
Db 71312 TGTGCAGAGAAAGAAATCCAAAGTAGGCTGCAGAGCAATCATTTGCTAG-----AGAGTTAG 71367

Qy 564 CATATGTAATGAGGAGCTGAATGTTAATCCTCAAGACATGGGAAAAATATCTCTCGAC 623
Db 71368 CATGACTAAAAGGGAGCAAGTGTCTAATATTCAGACAAATGTTAAAAAGGCTTGGAGGC 71427

Qy 624 ATGTCAGAGGTCCTCACAGCAGTCCATCAATCACTGCTGCTGGAGCCCTAG-----GA 676
Db 71428 ATTTTCAGAGATCTATGAAGCAGCCCTCCCATCAGAGTGCAGAGTTTGGTGGCACTAGG 71487

Qy 677 GAAATGTTTTTGGGACAGGCCAGGTCCTGCTGTGTCAGAGCTTAGAGACTTTGG 736
Db 71488 CCCAGAGTTTTATGGCCANNCCAGGCCACACTGCTATGCACAGCTTTGGGACACTG 71547

Qy 737 TGCCCTGTGTCCCACTTAA----- 755
Db 71548 CTGCCCGCATCCAGGCCACTCTGCTGCTCCACCTTGGCTCAAAAGGGCCAAAGATAG 71607

Qy 756 ---TTCAGCTGTGGCTTCAGAGGGTGCAGAGCCCAAGCCCTTGGCAGCTTCCAAAGTGGTGT 812
Db 71608 AGCTTGGACCACTGCTCCGAGGGCACAAGCCATGAAGCCTTGGTGGTTTCCATGTTGGTGT 71667

Qy 813 TGAGCTGTGGGTGCAAAAGAGTCAAGAAAT--TGAGGTTTGGGAACCTCCCAATCAGATTT 870
Db 71668 TAAGCTGCAGGTGCCAGAAATGCAGATTTGAGGGAGCTTGGGCACTTCCACCTAAATTT 71727

Qy 871 CAGAAGATATATGGAACCCCTGGATGCCAGGCAAGAGTTTGTGTAGGGGTGGGGTCC 930
Db 71728 CAGAGGATGTCTCAAGAACCCCTAGGTTCCAGGCAAGAGCATGATACAGGGGAGAGGCC 71787

Qy 931 TCATGGAGAACTCTGCAAGGGTAGTACAAAAGGAAATTTGGTGGGAGGCCCCACAC 990
Db 71788 TTGCAGAGAACTCTACTAGGGCAATGCCAAAGAAAATTTGGGGTTGGAGTCTCTCACAC 71847

Qy 991 AGAGTCCCCAGTGGGGCTCCATCTAGTAGAGTGTGAGAAGAGTCCACCATCTCTCCAGA 1050
Db 71848 ATGGTCCCCACTGGGGCACTACCTGGTGATACTGTGGGAATGGGGCTGCTGCCCCCTCCAGA 71907

Qy 1051 CTCAGAGAGGTTAGATCCACTGCAGACTTGCAGCATGTGCTCTGAAAAATTCACAGACACT 1110
Db 71908 CCCAGAGATGTAGATGCATGGCAGCTGGCACCTGAGCCCTGGAAGAGCTGCAGGCAC 71967

Qy 1111 CAGTGGCCAGCTGTGAAAGCAGCAGGATGGAGTGTGTACCTTACAAAACCGTAGTGCA 1170
Db 71968 CAACTCCAACCCATGAGATCAGCCACATGGGGTAC--TCCCAGGGAAGGCCACAGAGGCA 72025
```


Db 63617 GCTCCAGAAAGCAAGCCCTTGGCAGTGTCCACCTGGTGCTAATCTGCAGGG 63558
Qy 826 GCAAGAAGTCAGAATTTAGGTTTGGGAA---CCTCAATCAGATTTTCAAGATATAT 882
Db 63557 ATACAGAATGCAAGAGTTGTAGGCCCACTATGGCTTCACCTAGACTTCAGAGGATGTAT 63498
Qy 883 GGAACCCCTGGATGCCAGCAGCAAGTTTCTCTAGGGTGGGGTCTCATGGAGAAC 942
Db 63497 CTAAGCATGGGACCCAGCAGAGACTTGTCTACAGGAGCAGCCATGGCTGAGAGCC 63438
Qy 943 TC 944
Db 63437 CC 63436
RESULT 10
US-08-232-463-14/C
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232,463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935,313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: pTZgpt-Fls
US-08-232-463-14
Query Match 3.5%; Score 48.8; DB 1; Length 7218;
Best Local Similarity 4.9%; Pred. NO. 5.7e-05;
Matches 17; Conservative 192; Mismatches 139; Indels 0; Gaps 0;

Qy 358 TGGCTAATGAGGAACAAATTTTGGCTAATAAGGAAGAAATTTCTAAGCAGCAAGCAAT 417
Db 1462 TAGTTAAAGATAGAAGAATTTGGTACRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1403

Qy 418 CAAGAGGTGACTTGGTGCTGTTAAAGGCATTCAGTTTCATTAAGGAGCAGACATAG 477
Db 1402 RRR 1343
Qy 478 AGTTTCAGAAAATTTGCACCCCTGACAATGTGATAAAAAAGAAAACCCATTTTCTCAGGG 537
Db 1342 RRR 1283
Qy 538 AAATTCAGCTGGCTGCGCAAAATTTGCATATGTAATGAGGAGCTGAATGTTAATCCTCAA 597
Db 1282 RRR 1223
Qy 598 GACAATGGGAAATATCTCTCGCATGTGAGAGTCTTACACAGCTCATCAATCA 657
Db 1222 RRR 1163
Qy 658 CTGGCTGGAGGCTAGGAGAAATGTTTGTGGGAGCAGGCCAGGG 705
Db 1162 RRR 1115
RESULT 11
US-08-463-911-6
; Sequence 6, Application US/08463911
; Patent No. 5869330
; GENERAL INFORMATION:
; APPLICANT: Scherer, Philipp E.
; APPLICANT: Lodish, Harvey F.
; TITLE OF INVENTION: A NOVEL SERUM PROTEIN PRODUCED
; TITLE OF INVENTION: EXCLUSIVELY IN ADIPOCYTES
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
; STREET: Two Militia Drive
; CITY: Lexington
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02173
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/463,911
; FILING DATE:
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Granahan, Patricia
; REGISTRATION NUMBER: 32,227
; REFERENCE/DOCKET NUMBER: WHI95-05
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 861-6240
; TELEFAX: (617) 861-9540
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1313 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 73...804
US-08-463-911-6
Query Match 3.5%; Score 48.4; DB 2; Length 1313;
Best Local Similarity 48.8%; Pred. NO. 2.7e-05;
Matches 164; Conservative 0; Mismatches 166; Indels 6; Gaps 1;

Qy 41 ATGTAGAAAGTTGCTGGACCTCCAGCACACCCAGGCCCCAGAGAAAGTGGGCTCTG 100
Db 281 ATCCAGGTCTTATTGGTCTTAAGGAGAGATCGGTGAACCGGAGTACCCGGGCTGAG 340


```
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 861-6240
; TELEFAX: (617) 861-9540
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1276 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 46..786
; US-08-463-911-1

Query Match          2.9%; Score 40; DB 2; Length 1276;
Best Local Similarity 49.1%; Pred. No. 0.013;
Matches 106; Conservative 0; Mismatches 110; Indels 0; Gaps 0;

Qy 161 CTGATATAGAAAGTCAAGCTTTACTGTGAAGCTCAGTGGAAAGCTTCCTCTTCTTTCA 220
    ||||| | ||||| | ||||| | ||||| | ||||| | ||||| | ||||| |
Db 377 CTTATATGATCGCTCAGCGTTTCAGTGTGGGCTGGAGACCGCGTCACTGTCCCAATG 436

Qy 221 AGCCCATCATCTTCACAGGGTCTCTGTACATGCCCAGAGGGATTTAAAGGAGGCCATGG 280
    ||||| | ||||| | ||||| | ||||| | ||||| | ||||| | ||||| |
Db 437 TACCCATTCCGTTTACTAAGATCTTCTACAAACCAACAGAAATCATATGACGGCAGCACTG 496

Qy 281 GAGTCTTTGCTTGAGGGTGCCTGGGAATTACTACTCCAGCTTTCATGTTGAGCTGCATC 340
    ||||| | ||||| | ||||| | ||||| | ||||| | ||||| | ||||| |
Db 497 GCAAGTTCTACTGCAACATCCGGGACTCTACTCTCTTACCACATCAGCGGTGTACA 556

Qy 341 ATTGCAAGGTGAATATTTGGCTAATAGGAAGCAAA 376
    ||||| | ||||| | ||||| | ||||| | ||||| | ||||| | ||||| |
Db 557 TGAAGATGTCAAGTGAGCCTCTTCAGAAGGACA 592
```

RESULT 15
US-08-543-246B-8
; Sequence 8, Application US/08543246B
; Patent No. 6262244
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: DNA and amino acid sequence specific for
; TITLE OF INVENTION: natural killer cells
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Michael W. Glynn
; ADDRESSEE: No. 6262244artis Corporation
; STREET: 564 Morris Avenue
; CITY: Summit,
; STATE: NJ
; COUNTRY: US
; ZIP: 07901-1027
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA: US/08/543.246B
; APPLICATION NUMBER: US/08/543.246B
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/676.663
; FILING DATE: 28-MAR-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US92/02469
; FILING DATE: 27-MAR-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/122.514
; FILING DATE: 24-SEP-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Kassenoff, Melvyn M.
; REGISTRATION NUMBER: 26,389

```
; REFERENCE/DOCKET NUMBER: 118-7704/PCT/CONT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 908-522-6927
; TELEFAX: 908-522-6955
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1755 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 339..986
; FEATURE:
; NAME/KEY: mat_peptide
; LOCATION: 339..986
; US-08-543-246B-8

Query Match          2.8%; Score 38; DB 4; Length 1755;
Best Local Similarity 62.8%; Pred. No. 0.07;
Matches 59; Conservative 0; Mismatches 35; Indels 0; Gaps 0;

Qy 1283 TTCGACATTATATGGGCGCCGTACCCCTTTGTTTGGCCAAATTTTTCATTGGAACGTG 1342
    ||||| | ||||| | ||||| | ||||| | ||||| | ||||| | ||||| |
Db 1564 TTTGTACTTGCTTGAGTCCCATGACTGTTCTCTTCCTCTCTTCTCTCTTTTGAATAG 1623

Qy 1343 CCGTATTTACCAATGCGCTGACCTCCCATGTGAT 1376
    ||||| | ||||| | ||||| | ||||| | ||||| | ||||| | ||||| |
... Db 1624 TAATATCCATCCTATGTTTGTGCCCACTATTGTAT 1657
```

Search completed: February 19, 2003, 23:16:17
Job time : 910.904 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 19, 2003, 15:08:07 ; Search time 476.571 Seconds
(without alignments)
6506.903 Million cell updates/sec

Title: us-09-997-610-1-copy_2_1378

Perfect score: 1377

Sequence: 1 atagtgtcatacctgtctt.....gcctgtacctccattgtatg 1377

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N_Geneseq_101002.*

- 1: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*
- 2: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*
- 3: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*
- 4: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*
- 5: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*
- 6: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*
- 7: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*
- 8: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*
- 9: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*
- 10: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*
- 11: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
- 12: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*
- 13: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*
- 14: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*
- 15: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*
- 16: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*
- 17: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*
- 18: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*
- 19: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*
- 20: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*
- 21: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*
- 22: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.*
- 23: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.*
- 24: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1281	93.0	1338	24	AAI44066 Human genset metab
2	663	48.1	2590	22	AAH14327 Human cDNA sequenc
3	643.8	46.8	109201	24	ABQ88125 Human osteoblast d
4	640	46.5	155074	24	ABN85735 Human genomic regi
5	639.8	46.5	3093	22	AA503687 Rhesus gene locus:
6	639.8	46.5	9236	22	AA503689 Rhesus gene locus:
7	639.8	46.5	9236	22	AA503690 Rhesus gene locus:
8	639.8	46.5	9241	22	AA503688 Rhesus gene locus:
9	631	45.8	98690	24	ABX12169 Human DNA represen

10	618.4	44.9	6063	22	AAI03635 Human reproductive
11	618.2	44.9	465237	24	ABQ87681 Human oestrogen re
12	618.2	44.9	465237	24	ABA90193 Human oestrogen re
13	617.4	44.8	3203	22	AAI05134 Human secreted pro
14	616.8	44.8	6063	22	AAI03634 Human reproductive
15	602.8	43.8	1736	22	AAH13678 Human cDNA sequenc
16	595.6	43.3	22509	22	AA544505 Human LEKTI DNA cl
17	595.6	43.3	31529	22	AA544505 Human LEKTI DNA cl
18	590.8	42.9	1278	22	AAI05178 Human secreted pro
19	590	42.8	32127	22	AAI92555 Human secretory re
20	590	42.8	32127	22	AAI92555 Human kidney relat
21	590	42.8	72215	22	AAK86832 Human immune/haema
22	580.4	42.1	1485	23	AA591984 DNA encoding novel
23	563.8	40.9	8387	22	ABAI4892 Human nervous syst
24	563.8	40.9	8387	22	AAK84548 Human immune/haema
25	559.2	40.5	7726	22	ABA21079 Human nervous syst
26	558.2	40.5	1580	23	AA587262 DNA encoding novel
27	557	40.5	159400	24	ABQ88126 Human osteoblast d
28	556.6	40.4	1621	23	AA567924 DNA encoding novel
29	555.8	40.4	81369	21	AA597997 Human T gene DNA.
30	551.6	40.1	1306	23	AA564678 DNA encoding novel
31	549	39.9	3049	22	AAH16637 Human cDNA sequenc
32	543	39.4	1405	22	AAK70265 Human immune/haema
33	539.8	39.2	1405	22	AAK70265 Human immune/haema
34	523	38.0	1946	22	ABA44227 Human breast cell
35	523	38.0	1946	22	ABA54677 Human foetal liver
36	523	38.0	1946	22	ABA24459 Probe #2925 for ge
37	523	38.0	1946	22	AAK02964 Human brain expres
38	523	38.0	1946	22	AAK28410 Human bone marrow
39	523	38.0	1946	22	AAI12975 Probe #2908 for ge
40	523	38.0	1946	22	AAI34334 Probe #3020 used t
41	523	38.0	1946	22	AAI02893 Probe #2884 used t
42	523	38.0	1946	24	ABS02918 Human genome-deriv
43	519.4	37.7	22756	22	AA540321 DNA encoding human
44	519.4	37.7	22756	22	AAI03921 Human reproductive
45	515.2	37.4	923	23	AA566623 DNA encoding novel

ALIGNMENTS

RESULT 1

AAI44066

ID AAI44066 standard; cDNA: 1338 BP.

XX AAI44066;

AC AAI44066;

XX AAI44066;

DT 27-SEP-2002 (first entry)

DE Human genset metabolic gene (GMG-9) cDNA sequence.

XX Human genset metabolic gene (GMG-9) cDNA sequence.

KW Human; gene; ss; gene therapy; genset metabolic gene; GMG-7A; GMG-7B;
KW GMG-8; GMG-9; GMG-10; GMG-11; metabolic-related disorder; obesity;
KW impaired glucose tolerance; insulin resistance; Syndrome X;
KW Type II diabetes; hyperlipidaemia; atherosclerosis; hypertension;
KW heart disease; cardiac insufficiency; coronary insufficiency;
KW high blood pressure; insulin sensitizer;
KW non-insulin dependent diabetes mellitus.

XX Homo sapiens.

OS Homo sapiens.

Key Location/Qualifiers

CDS 1..1338

FT /*tag= a

FT /partial

FT /product= "Human GMG-9 protein"

FT /note= "No stop codon is given"

XX WO200255694-A2.

XX 18-JUL-2002.

XX 15-JAN-2002; 2002WO-IB01215.

XX 15-JAN-2002; 2002WO-IB01215.

XX 16-JAN-2001; 2001US-262235P.
PR (GEST) GENSET.
XX
XX Erickson MR, Bour BA, Bihain B, Tanaka H;
PA
XX WPI; 2002-557821/59.
DR P-PSDB; AAO15423.
DR
XX Treating or preventing a metabolic-related disease or disorder, e.g.
PT obesity, impaired glucose tolerance, insulin resistance, Syndrome X, or
PT Type II diabetes, comprises administering Genset Metabolic Genes -
XX
XX Disclosure; Page 122-124; 128pp; English.
XX
XX The invention comprises the amino acid and coding sequences of six human
CC genset metabolic genes (GMG-7A, GMG-7B, GMG-8, GMG-9, GMG-10 and GMG-11).
CC The GMG DNA and protein sequences of the invention are useful for
CC treating or preventing metabolic-related disorders, such as: obesity;
CC impaired glucose tolerance; insulin resistance; Syndrome X; Type II
CC diabetes; hyperlipidaemia; atherosclerosis; hypertension; and heart
CC diseases (e.g. cardiac insufficiency, coronary insufficiency or high
CC blood pressure). The GMG DNA and protein sequences of the invention may
CC also be used as insulin sensitisers - for improving insulin sensitivity
CC in persons with non-insulin dependent diabetes mellitus. The present cDNA
CC sequence encodes the human GMG-9 protein.
XX
XX Sequence 1338 BP; 359 A; 288 C; 350 G; 341 T; 0 other;
SQ
Query Match 93.0%; Score 1281; DB 24; Length 1338;
Best Local Similarity 99.6%; Pred. No. 0;
Matches 1284; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
Oy 89 TGGGGCTCCTGGTGCACAGGTTTACACAAATATACAGAGAAATAGTGAATGACAA 148
Db III III III III III III III III III III III III III III III III III
Oy 149 AATGCCCTGCTCATATAGAAAGTTCAGCCTTTACTGTGAGCTCAGTGGAAACATC 208
Db III III III III III III III III III III III III III III III III III
Oy 209 CTCCTCCCTTCAAGCCCATCATCTTTCACAGGGTGCCTGTACAAATGCCAGAGGATTTAA 268
Db 170 CTCCTCCCTTCAAGCCCATCATCTTTCACAGGGTGCCTGTACAAATGCCAGAGGATTTAA 229
Oy 269 AGGAGCCATGGAGTCTTTTGTGTCAGGGTGCCTGGGAATTTACTTCCAGCTTTGTATG 328
Db 230 AGGAGCCATGGAGTCTTTTGTGTCAGGGTGCCTGGGAATTTACTTCCAGCTTTGTATG 289
Oy 329 TTGAGCTGCATCATTCGAAAGTGAATATTTGGCTAATGAGGAAGCAAAATTTTGGCTAATA 388
Db 290 TTGAGCTGCATCATTCGAAAGTGAATATTTGGCTAATGAGGAAGCAAAATTTTGGCTAATA 349
Oy 389 AGGAAGAAATTTCTAAGCAGCAAGCATTTCAAGAGGTGACTTTGGGTGCTGTTTAAAGGCAT 448
Db 350 AGGAAGAAATTTCTAAGCAGCAAGCATTTCAAGAGGTGACTTTGGGTGCTGTTTAAAGGCAT 409
Oy 449 TCAGTTTTCATAAGGGAGGAGCAGCATAAGAGTTTCAGAAATTTGCACCCCTGACAAATGTGA 508
Db 410 TCAGTTTTCATAAGGGAGGAGCAGCATAAGAGTTTCAGAAATTTGCACCCCTGACAAATGTGA 469
Oy 509 TAAAAAGAAAACCCATTTTCTGAGGGAAATTTCAAGCTGGCTGCAGAAATTTGCATAT 568
Db 470 TAAAAAGAAAACCCATTTTCTGAGGGAAATTTCAAGCTGGCTGCAGAAATTTGCATAT 529
Oy 569 GTAATGAGGAGCTGAATTTAATCTCTCAAGACAATGGGAAATATCTCTGGACATGTC 628
Db 530 GTAATGAGGAGCTGAATTTAATCTCTCAAGACAATGGGAAATATCTCTGGACATGTC 589
Oy 629 AGAGGTCTTCACAGCAGTCCATCAATCACTGGCCTGGAGGCTTAGGAGAAATGTTTT 688
Db 590 AGAGGTCTTCACAGCAGTCCATCAATCACTGGCCTGGAGGCTTAGGAGAAATGTTTT 649

Oy 689 GTGGGACAGGCCAGGGTCCCTGTGTGTGTCAGCTAGAGACTTGGTGCCTGTGTGTC 748
Db III III III III III III III III III III III III III III III III III
Oy 749 CAGTTTAAATTCAGCTGTGGCTTCAGAGGTGCAAGCCCCAAGCCTTGGCAGCTTCCCAAGTG 808
Db III III III III III III III III III III III III III III III III III
Oy 809 GTGTTGAGCCTGTGGTGCAGAAAGTCAAGAAATTCAGGTTTGGGAACCTCCCAATCAGAT 868
Db 770 GTGTTGAGCCTGTGGTGCAGAAAGTCAAGAAATTCAGGTTTGGGAACCTCCCAATCAGAT 829
Oy 869 TTCAGAAGATATATGAAACCCCTTGATGCCAGGCAGAAATTTGCTGTAGGGGTGGGT 928
Db 830 TTCAGAAGATATATGAAACCCCTTGATGCCAGGCAGAAATTTGCTGTAGGGGTGGGT 889
Oy 929 CCTCATGGAAGAACCTCTGCAAGGGTAGTACAAAGGAAATTTGGGTGGGAGCCCCAC 988
Db 890 CCTCATGGAAGAACCTCTGCAAGGGTAGTACAAAGGAAATTTGGGTGGGAGCCCCAC 949
Oy 989 ACAGAGTCCCCAGTGGGGTCCATCTAGTAGAGCTGTGAGAGAGTCCACCATCCTCCA 1048
Db 950 ACAGAGTCCCCAGTGGGGTCCATCTAGTAGAGCTGTGAGAGAGTCCACCATCCTCCA 1009
Oy 1049 GACTCCAGAAGGTAGATCCACTGCAGCCTTGCAGCATGTGCTGAAAAATCCACAGACA 1108
Db 1010 GACTCCAGAAGGTAGATCCACTGCAGCCTTGCAGCATGTGCTGAAAAATCCACAGACA 1069
Oy 1109 CTCAGTGCCAGCCTGTGAAAGCAGCAGGGATGGAGTCTGTACCTACAAAACCGTAGTGG 1168
Db 1070 CTCAGTGCCAGCCTGTGAAAGCAGCAGGGATGGAGTCTGTACCTACAAAACCGTAGTGG 1129
Oy 1169 CAGAGCTGACCAAGACCGTGGGAATCTACTCTTGCAATTGTATGACCTGGAGCTGAGAC 1228
Db 1130 CAGAGCTGACCAAGACCGTGGGAATCTACTCTTGCAATTGTATGACCTGGAGCTGAGAC 1189
Oy 1229 ATGGAGTCAAAAGAGATCATTTTGGAGCTTTAAGATTTCACCTGCCCACTGGATTTCGGA 1288
Db 1190 ATGGAGTCAAAAGAGATCATTTTGGAGCTTTAAGATTTCACCTGCCCACTGGATTTCGGA 1249
Oy 1289 CTTATATGGGGCCGTACCCCTTTTGTGTTTGGCAATTTTTCATTGGAACTGCCGTAT 1348
Db 1250 CTTATATGGGGCCGTACCCCTTTTGTGTTTGGCAATTTTTCATTGGAACTGCCGTAT 1309
Oy 1349 TTACCCAATGCCTGTACCTCCATTTGTATG 1377
Db 1310 TTACCCAATGCCTGTACCTCCATTTGTATG 1338
RESULT 2
AAH14327
ID AAH14327 standard; cDNA; 2590 BP.
XX
XX AAH14327;
XX
XX 26-JUN-2001 (first entry)
DT
XX
DE Human cDNA sequence SEQ ID NO:11697.
XX
KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX
XX Homo sapiens.
OS
PN EP1074617-A2.
XX
XX 07-FEB-2001.
PD
XX
PF 28-JUL-2000; 2000EP-0116126.
XX
PR 29-JUL-1999; 99JP-0248036.
PR 27-AUG-1999; 99JP-0300253.
PR 11-JAN-2000; 2000JP-0118776.

02-MAY-2000; 2000JP-0183767.
09-JUN-2000; 2000JP-0241899.
(HELI-) HELIX RES INST.

Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
WPI; 2001-318749/34.

Primer sets for synthesizing polynucleotides, particularly the 5602
full-length cDNAs defined in the specification, and for the detection
and/or diagnosis of the abnormality of the proteins encoded by the
full-length cDNAs -

Claim 8; SEQ ID 11697; 2537pp + CD ROM; English.

The present invention describes primer sets for synthesizing 5602
full-length cDNAs defined in the specification. Where a primer set
comprises: (a) an oligo-dr primer and an oligonucleotide complementary
to the complementary strand of a polynucleotide which comprises one of
the 5602 nucleotide sequences defined in the specification, where the
oligonucleotide comprises at least 15 nucleotides; or (b) a combination
of an oligonucleotide comprising a sequence complementary to the
complementary strand of a polynucleotide which comprises a 5'-end
sequence and an oligonucleotide comprising a sequence complementary to a
polynucleotide which comprises a 3'-end sequence, where the
oligonucleotide comprises at least 15 nucleotides and the combination of
the 5'-end sequence/3'-end sequence is selected from those defined in
the specification. The primer sets can be used in antisense therapy and
in gene therapy. The primers are useful for synthesizing polynucleotides
particularly full-length cDNAs. The primers are also useful for the
detection and/or diagnosis of the abnormality of the proteins encoded by
the full-length cDNAs. The primers allow obtaining of the full-length
cDNAs easily without any specialised methods. AH03166 to AH03628 and
AH03633 to AH03742 represent human cDNA sequences; AAB92446 to
AAB95893 represent human amino acid sequences; and AH03629 to AH03632
represent oligonucleotides, all of which are used in the exemplification
of the present invention.

Sequence 2590 BP; 722 A; 526 C; 670 G; 672 T; 0 other;

Query Match 48.1%; Score 663; DB 22; Length 2590;
Best Local Similarity 80.7%; Pred. No. 7.5e-197;
Matches 826; Conservative 0; Mismatches 160; Indels 37; Gaps

QY 391 GAAGAAATTTCTAAGCAGCAAAACCATTCAGAGGTCACITGGGTGCTGTTAAAGGCATTC 450
|||||
Db 609 GAAGAAATTTCTAAGCAGCAAAACCATTCAGAGGTCAGTGGGTGCTGTTAAAGGCATTC 668
|||||
QY 451 AGTTTCAATAGGGAGGCAGAGCATTAAGAGTTTCAGAAAAATTCACCCCTCACAAATGTGATA 510
|||||
Db 569 AGTTTCAATAGGGAGGCAGAGCATTAAGAGTTTCAGAAAAATTTGAGCCTGACATGCAATA 728
|||||
QY 511 AAAAAGAAAAACCCATTTTCTGAGGGGAAATTCACAGTCGGCTGCAGAAAATTTTCATATGT 570
|||||
Db 729 GAAAAGAAAAACCCAGTTTTCTGGGAGAAATTCACGCCAGCTCCAGAAATCTGCATTAAC 788
|||||
QY 571 AATGAGGAGCTGAATGTTAATCTCAAGACAATGGGGAAAAATATCTCCGACATGTCTAG 630
|||||
Db 789 ATCAAGAGGCTAATGTTAATCCCGAGGACCATGGGGAAAAATGCTCCAGGCCATGTCTAG 848
|||||
QY 631 AGGCTTTTCAGCAGCAGTCCATCAATCACTCGCCCTGGAGCCCTAGGAG --- AAAATGGTTT 687
|||||
Db 849 AGACCTTTCACAGCAGCCCCCTCTCTGTACAGGCCACAGATCCAGGAGGAAAAAATGGTTT 908
|||||
QY 688 TGTGGGAGGCCCGCAGGGTCCCTGTGCTGTGTGCAGCCTAGACATTTGCTGCCCTGTGTCT 747
|||||
Db 909 TATGGCCAGGACCAGGGTCCCTGTGCTGTGTGCAGCCTATAGACTTGGTGGCCCTGTGTC 968
|||||
QY 748 CCAGTTTAATTCA-----GCTGTGCTTTCAGAG 774
|||||
Db 969 CCAGCTGCTCCAGGCATGGCTGAAAGGGGCCCAATGTAGAGCTCATGCTGTGCTTTCAGAG 1028
|||||

PI Ji D, Axelrod DW, Cook JS, Jaiswal N, Einstein R, Houghton A;
PI Mertz L;
XX
DR WPI; 2002-557663/59.
XX
PT Use of genes and their expression profiles associated with osteoblast
PT differentiation for screening modulators bone formation, for diagnosing
PT or treating e.g. osteoporosis, or as markers for the differentiation
PT process -
XX
XX
PS Claim 1; SEQ ID NO 32; 78pp + Sequence Listing; English.
XX
CC The invention relates to genes and their expression profiles are used
CC for:
CC (a) screening modulators of precursor stem cell differentiation into
CC osteoblasts, or bone tissue deposition;
CC (b) diagnosing abnormal deposition of bone tissue, abnormal rate of
CC osteoblast formation or osteoporosis; or
CC (c) treating or monitoring treatment of the conditions cited in (b), or
CC monitoring the progression of bone tissue deposition.
CC Specific conditions include postmenopausal osteoporosis, glucocorticoid
CC osteoporosis or male osteoporosis, osteopenia, osteodystrophy,
CC drug-induced abnormalities in bone formation or bone loss, conditions
CC that involve altered bone metabolism (e.g. idiopathic juvenile
CC osteoporosis), skeletal disease linked to breast cancer, mastocytosis,
CC Fanconi syndrome or fibrous dysplasia. The present sequence is that of an
CC osteoblast differentiation associated cDNA marker of the invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 109201 BP; 32871 A; 23488 C; 22108 G; 30734 T; 0 other;
Query Match 46.8%; Score 643.8; DB 24; Length 109201;
Best Local Similarity 79.3%; Pred. No. 7.3e-190;
Matches 843; Conservative 0; Mismatches 177; Indels 43; Gaps 5;
QY 351 GAATAATTTGGCTAATGAGGAACAATTTTGGCTAATAAGGAAGAAATTTCTAAGCAGCA 410
DB 22659 GAACATTGAACCTTGAGAGAGATGATTTAGGGTATCTGGCAGAGAATAATTTCTAAGCAGCA 22600
QY 411 AAGCAATCAAGAGGTGACTTGGGTGCTGTAAAGGCAATTCAGTTTCATAAGGGAGGCAGA 470
DB 22599 AAGCAATTAAGAGATGACTTGGGTGCTGTAAAGCAATTTATTTTATAAGGGAAGCAGA 22540
QY 471 GCATAAGAGTTCAGAAAATTTCCACCCTGCACAAATGTGATAAAGAAAAGAAACCCATTTTC 530
DB 22539 GCATAMAGTTTGGAAATTTGCATCTCTGACAAATGTGATAAAGAAAGACAATCCTATTTTC 22480
QY 531 TGAGGGAAATTCAGCTGGCTGCAGAAATTTGCATATGTAATGAGGAGCTGAATGTTAA 590
DB 22479 TGAGGAGACATTCAGTCAGCTGCAGAAATTTGCATAAGTAATGAGGAGCTTAATTTAA 22420
QY 591 TCCTCAAGACAATGGGAAAATATCTCTCGACATGTGCAGAGTCTTCACAGCAGTCCAT 650
DB 22419 TCCCAAGACAATGGGAAAATGTCTCCAGGCGCATGTCAGAAAGTCTTCACAGCAGCCCT 22360
QY 651 CAATTCACGTGCTGGAGGCCTA---GGAGAAAATGGTTTCTGGCAGACAGGCCAGGGTC 707
DB 22359 CCAATCAGAGGCTGGAAGCCTAGAGGAAAAGTGTTTTTTGGGCCAGGCCAGGGTC 22300
QY 708 CCTGTGCTGTGCAGCCTTAGAGACTTGGTGCCTGTGTGCCAGTTAATTCAGCTGTGCG 767
DB 22299 CTCATGCTGTGCAGCCTTAGAGACTTGTGCCCTGCATCCAGCCACTGCAGCTATGGG 22240
QY 768 T-----TCAGAGGGTGCAGCCCAAGCCCTT 793
DB 22239 TAAAAAGGGCCCAACACAGAGCTCAACAACACAGAGCTTGAAGGGTGCAGAGCTCAAGCTTT 22180
QY 794 GCAGCTTCCAAGTGTGTGAGCTGTGGGTGCAAGAGCTCAACAATTCAGGTTGGG 853
DB 22179 GGTAGCTCCATGTGTGTGATCTGCGAGTGCACAGAAAGCAAGATTTGGGGTTGGG 22120

QY 854 AACCTCCATCAGATTTTCAGAAAGATATATGAAACCCCTGGATGCCAGCGAGAAAGTTTG 913
DB 22119 AACTTGTGCTAGATTTC--AGATCTATGAAATGCTTTGGATGTCAGCGAGAAAGTTTG 22063
QY 914 CTGTAGGGTGGGGTCTCATCGAGAAACCTCTGCAAGGAGTGTACAAAAGGAGAAATGTTG 973
DB 22062 CTATGAG--GGGGTCTCTGTAATAATCTGCTAGGCGAGTGCAGAAAGAAATGTGG 22005
QY 974 GTTGGAGGCCCCACACAGAGTCCCGAGTGGGCTCCATCTAGTAGAGCTGTGAGAAGAA 1033
DB 22004 GGTAGGAGCCCCCATACAGAGTCTCTACTGGGCAACCACTAGTGGAGCTGTGAGAAAAG 21945
QY 1034 GTCCACCATCTCCAGACTCCAGAGGGTAGATCCACTGACAGCTTTGCAGCATGTGCCCTG 1093
DB 21944 GCCCACCATCTCCAGAGCCCAAGATGATGATGCTCCACTGACAGCTTATGTCCTG 21885
QY 1094 AAAATCCACACACTCAGTCCAGCCCTGTGAAGACACAGGAGGTGAGTCTGTACCCT 1153
DB 21884 GAAAAGCCATAGACACTCAACACAGGCTGTGAAGACGCCAGGAGGAGGCTGTACCCT 21825
QY 1154 ACAAAACCCGTAGTGGCAGAGCTGACCAAGACCGTGGGAATCTACCTCTTTCATTTGCATG 1213
DB 21824 GCAAAGCCACAGGGTGAAGCTGACCAAGACTATGGGAACCTACTTCTTGCATCAGTGTG 21765
QY 1214 ACCTGAGCTGAGACATGAGTCAAAAGAGATCAATTTTGGAGCTTTAAGATTTGACTGCC 1273
DB 21764 ACCTGGATGTGAGACATGCTGTCAAAAGAGATCCTTTTGGAGCTTTAAGATTTGACTGCC 21705
QY 1274 CCATGAGATTTGGAGCTTATATGGGCCCGTA--CCCTTTTGGGCCAATTTTTCCTCA 1332
DB 21704 CTGTAGATTTTCAGACTTGAATGGGCCCTGTAGCCCTTTTGTGGGCCAATTTTTCCTCA 21645
QY 1333 TTTGGAACCTGCGTATTTTACCCTGCTACCTCCATTTGTA 1375
DB 21644 TTTGGAATGCTGATTTACTCAATGCTGCATCCCTATTTGTA 21602
RESULT 4
ABN85735/C
ID ABN85735 standard; DNA; 155074 BP.
XX
AC ABN85735;
XX
DT 30-SEP-2002 (first entry)
DE Human genomic region containing the ltrpc5 gene SEQ ID NO 6.
XX
DE Human; ltrpc5; taste; cell signalling; TC-ICS; food; pharmaceutical;
KW taste cell-specific ion channel subunit; chromosome 11p15-5; gene; ds.
XX Homo sapiens.
OS
XX WO200254069-A1.
PN
PD 11-JUL-2002.
XX
PF 26-DEC-2001; 2001WO-US49808.
XX
PR 29-DEC-2000; 2000US-259379P.
PR 21-DEC-2001; 2001US-0026188.
XX
XX (REGC) UNIV CALIFORNIA.
PA
XX Zuker CS, Zhang Y;
XX
DR WPI; 2002-583632/62.
XX
PT Identifying modulators of taste signaling in taste cells for use in
PT food and pharmaceutical industries to customize and regulate taste, by
PT determining effect of the compound on a taste cell-specific ion channel
PT subunit -
XX
PS Disclosure: Page 212-301; 306pp; English.

XX The invention relates to identifying (MI) a compound that modulates taste
CC signalling in taste cells, by contacting the compound with a eukaryotic
CC host cell or cell membrane which expresses a taste cell-specific ion
CC channel subunit (TC-ICS), and determining a functional effect of the
CC compound upon a transmembrane ion flux of a predetermined ion,
CC identifying a compound that modulates taste signalling in taste cells.
CC (MI) is useful for identifying a compound that modulates taste signalling
CC in taste cells, for identifying a compound that binds to a taste cell
CC specific ion channel subunit and for modulating taste signalling in taste
CC cells of a mammal, in particular a human. Modulators identified by (MI)
CC are used by the food and pharmaceutical industries to customize taste,
CC e.g. as additives to food or medicine so that the food or medicine tastes
CC different to the subject who ingests it. Bitter medicines can be made to
CC taste less bitter and sweet substance can be enhanced. The modulators are
CC useful for pharmacological and genetic modulation of taste signalling
CC pathways. The taste modulators can be directly administered to mammalian
CC subjects for modulation of taste in vivo. The present sequence is that of
CC the human genomic region (Chromosome 11p15-5) containing the *ltl* gene
XX of the invention.
SQ Sequence 155074 BP; 32795 A; 41240 C; 44622 G; 36402 T; 15 other;

Query Match 46.5%; Score 640; DB 24; Length 155074;
Best Local Similarity 80.6%; Pred. No. 1.4e-188;
Matches 825; Conservative 0; Mismatches 160; Indels 39; Gaps 5;

QY 390 GGAAGAAATTTCTAAGCAGCAAGCATTCAGAGGTGACTTGGGTGCTTTAAAGGCATT 449
DB 94900 GGAAGAAATTTCTAAGCTGCAAGAGCTTTCAAGAGGTGACTTGGGTGCTTTAAAGGCATT 94841
QY 450 CAGTTTCAATAGAGGAGGAGCAGCATAAGAGTTTCAGAAATTTGCACCTGCAATGTGAT 509
DB 94840 CAGTTTCAATAGAGGAGGAGCAGCATAAGAGTTTCAGAAATTTGCACCTGCAATGTGAT 94782
QY 510 AAAAAAGAAAACCCATTTCTGAGGGGAAATTCAGCTGGCTGCAGAAATTTGCATATG 569
DB 94781 AAAAAAGAAAACCCATTTCTGAGGGGAAATTCAGCTGGCTGCAGAAATTTGCATATG 94722
QY 570 TAATGAGGAGCTGAATGTTAATCTCAAGACAATGGGAAATATCTCTCGACATGCA 629
DB 94721 TAACNAGAAGCAATGTTAATCTCAAGACAATGGGAAATATCTCTCGACATGCA 94662
QY 630 GAGGTCTTACAGCAGTCCATCAATCACTGGCTGGAGGCTAGGAG---AAAATGTTT 686
DB 94661 GAGGTCTTACAGCAGTCCATCAATCACTGGCTGGAGGCTAGGAG---AAAATGTTT 94602
QY 687 TTGTGGGACAGCCAGGGTCCCTGCTGTGTGTCAGCCTAGAGACTTGGTGGCCCTGTGT 746
DB 94601 TCCTGGAGTGGGCCAGGGCCCTCCATGCTGTGTGTCAGCCTAGGAGTGGTGGCCCTGTGT 94542
QY 747 CCCAG-----TTAATTCAGCTGTGGCTTCAGA 773
DB 94541 CCCAGCCCTCTAGCCATAGCTATAAAGGTGTCAAGGTTCAGCTTGGGCTGTGGCTTCAGA 94482
QY 774 GGGTCAAGCCCAAGCCTTGGAGCTTCCAAAGTGTGTGAGCCTGTGGGTGCAAGAA 833
DB 94481 GGGTCAAGCCCAAGCCTTGGAGCTTCCAAAGTGTGTGAGCCTGTGGGTGCAAGAA 94422
QY 834 GTCAGAAATTCAGGTTTGGGAACCTCCAAATCAGATTTTCAGAGATATATGGAACCCCTG 893
DB 94421 GTCAGAAATTCAGGTTTGGGAACCTCCAAATCAGATTTTCAGAGATATATGGAACCCCTG 94363
QY 894 GATGCCAGGAGAGATTTGCTGTAGGGGTGGGTCCCTCATGGAGAACCTCTGCAAGGT 953
DB 94362 GATGCCAGGAGAGATTTGCTGTAGGGGTGGGTCCCTCATGGAGAACCTCTGCTAGGAG 94303
QY 954 AGTACAAAAGGAAATGTTGGGTGGAGCCGCCACACAGAGTCCCAAGTGGGGTCTCCATC 1013
DB 94302 AGTACAAAAGGAAATGTTGGGTGGAGCCGCCACACAGAGTCCCAAGTGGGGTCTCCATC 94243
QY 1014 TAGTAGAGCTGTGAGAGAAATGCCACCATCTCCAGATCTCCAGAGGAGGTAGATCCACTGA 1073
DB 94243 TAGTAGAGCTGTGAGAGAAATGCCACCATCTCCAGATCTCCAGAGGAGGTAGATCCACTGA 1073

DB 94242 TAGTGAGCTGTGAGAGAGAGGGCCCACTGTCTCCAGATCCCAGAAATGATAGATCCATCAA 94183
QY 1074 CAGCTTGGAGCATGTGCTGAAATAATCCACAGACACTCAGTGCCAGCCTGTGAAAGCAGC 1133
DB 94182 CAGCTTGGAGCATGTGCTGAAATAATCCACAGACACTCAGTGCCAGCCTGTGAAAGCAGC 94123
QY 1134 AGGATGAGCTGTACCTTACAAACCTGAGAGCTGAGAGCTGACCAAGACCGTGGGAAT 1193
DB 94122 CAGGAGGAGGCTGTACCTTACAAACCTGAGAGCTGAGAGCTGACCAAGACCTATGGGAAC 94063
QY 1194 CTACCTCTTGCAATGTGACCTGACCTGAGCTGAGAGCTGAGAGCTGAGAGCTGAGAGCTATTTGG 1253
DB 94062 CCACCTCTTGCAATGTGACCTGAGCTGAGAGCTGAGAGCTGAGAGCTGAGAGCTATTTGG 94003
QY 1254 AGCTTTAAGATTTGACTGCCCCACTGGATTTCCGGACTTATATGGGCCCGTA-CCCTCTTT 1312
DB 94002 AGCATTAAGATTTAAGTGGCCCTGAGTTCGAACTTTCGAACTTTCGAACTTTCGAACTTTC 93943
QY 1313 GTTTTGGCCAAATTTTTCATTTGGAACCTGCGGTATTTACCAATGCTGTACCTCCATT 1372
DB 93942 GTTTTGGCCAAATTTACTCCCATTTCAATGGGTGTATTTATCCAGGCTGTACCTCAAT 93883
QY 1373 GTAT 1376
DB 93882 GTGT 93879
RESULT 5
AAS03687/C
ID AAS03687 standard; DNA; 3093 BP.
XX AAS03687;
AC AAS03687;
XX AC AAS03687;
XX AC AAS03687;
DT 29-AUG-2001 (first entry)
XX 29-AUG-2001 (first entry)
DE Rhesus gene locus: RHD gene deletion in Rh negative haplotypes.
XX Rhesus box: RHD positive; sequence length polymorphism; SSP; RHD; SMP1;
KW RHE; Rh negative; blood group typing; blood transfusion; antlgen C;
KW haemolytic disease of the newborn; chromosome 1 p34.1-p36; ds.
XX Homo saplens.
OS Homo saplens.
XX Homo saplens.
FH Key Location/Qualifiers
FT primer_bind 32..54
FT /*tag= a
FT /note= "Binding site of primer rez7"
FT primer_bind complement (3034..3054)
FT /*tag= b
FT /note= "Binding site of primer rnb31"
XX WO200132702-A2.
PN WO200132702-A2.
XX WO200132702-A2.
PD 10-MAY-2001.
XX 10-MAY-2001.
XX 31-OCT-2000; 2000WO-EP10745.
XX 31-OCT-2000; 2000WO-EP10745.
PR 02-NOV-1999; 99EP-0121686.
PR 31-MAY-2000; 2000EP-0111696.
XX 31-MAY-2000; 2000EP-0111696.
XX (DRKB-) DRK BLUTSPENDEDIENST BADEN WUERTTEMBERG.
XX Flegel WA, Wagner FF;
PI Flegel WA, Wagner FF;
XX Flegel WA, Wagner FF;
DR WPI; 2001-291052/30.
XX WPI; 2001-291052/30.
XX New nucleic acid molecular structure, useful for detection of common
PT RHD positive haplotypes in D-negative individuals, comprises RHD, SMP1
PT and RHE genes -
XX RHD positive haplotypes in D-negative individuals, comprises RHD, SMP1
PS Example 10; Flg 5; 135pp; English.
XX Example 10; Flg 5; 135pp; English.
CC The sequence represents the coding sequence of Rhesus gene locus:

CC RHD gene deletion in Rh negative haplotypes. The Rhesus genes
CC locus comprises the RHD, SMPI and RHCE (all undefined) genes and/or the
CC Rhesus box(es), preferably the hybrid Rhesus box, the upstream Rhesus box
CC and/or the downstream Rhesus box. The RHD and RHCE genes are located at
CC chromosome 1 p34.1-p36. Rhesus box flanks the breakpoint region of the
CC RHD deletion in the common RHD negative haplotypes. The sequence has
CC been used to design primers which are useful for: (1) the specific
CC detection of the common RHD positive haplotypes in D-negative
CC individuals; (2) blood group typing; (3) determining whether a patient
CC can be transfused with RhD negative blood and whether blood is suitable
CC for transfusion to patients who should not be exposed to antigen C; (4)
CC assessing the risk of a RhD negative mother of conceiving or carrying an
CC RhD positive foetus. Anti-D antibodies are useful for treating pregnant
CC women who are Rhesus D negative, where the foetus is not homozygous for
CC the RHD gene to treat or prevent haemolytic disease of the newborn.
XX
SQ Sequence 3093 BP; 891 A; 754 C; 619 G; 829 T; 0 other;

Query Match 46.5%; Score 639.8; DB 22; Length 3093;
Best Local Similarity 79.1%; Pred. No. 1.6e-189;
Matches 841; Conservative 0; Mismatches 177; Indels 45; Gaps 5;

QY 351 GAATATTGGCTAATCAGAGCAAAATTTGGCTAATAGGAAGAAATTTCTAAGCAGCA 410
DB 1377 GAACTTTGAACCTTCAGAGAGATGATTTAGAGTATCTGCAGAGAGAAATTTCTAAGCAGCA 1318
QY 411 AAGCATTCAAGAGGTGACTTGGTGTCTTAAAGGCATTTCAGTTTCATAAAGGGAGGAGCA 470
DB 1317 AAGCATTCAAGAGGTGACTTGGTGTCTTAAAGACATTCAGTTTATAAGGGAAGCAGCA 1258
QY 471 GCATAGAGTTTCAGAAAAATTTGCACCCCTGACAAATGTGATATAAAAAAGAAAAA-CCCATTTT 528
DB 1257 GCATAAAAGTTCCGAAAAATTTGCAGCCTGCACAAATGTGATAGAAAAAGAAAAATTTCCCATTT 1198
QY 529 TCTGAGGGAAATTCAGCTGCTGCAGAAATTTGCATATGTAAATCAGGAGCTGAATGTT 588
DB 1197 TCTGAGGAGAAATTCAGCTGCTGCAGAAATTTGCATATGTAAATCAGGAGCTGAATGTT 1139
QY 589 AATCCTCAAGACAATGGGAAAAATATCTCTGGACATGTCAGAGTCTTCACAGCAGTCC 648
DB 1138 AATTCCAAGACANATGGGAAAAATGTCTCCAGGCGATGTCAGAGGTCTTTATGGCAACCC 1079
QY 649 ATCAATCACTGCGCTGGAGCGCT-AGGAGAAAAATGGTTTTTGGGACAGAGCCAGGGTC 707
DB 1078 CTCCTCATCAGGTCCAGAGGTATCAGGAAAAATGGTTTTTGTGGCCAGGCCCGGGGTC 1019
QY 708 CTTGTGCTGTGTGCAGCCTAGAGACTTGGTGCCTGTGTCCAGTTA-----754
DB 1018 CTCATGCTGTGTGCAGCCTTAGGAGCTTGGTGCCTGTGCATCCCGAGCCACTCCCAACCATGA 959
QY 755 -----ATTGAGCTGTGGCTTCAGAGGGTGCAGGCCCAAGCCCAAGCCTT 793
DB 958 CTGACGGGAGGCAAGGTAGAGCTTGGGCTGTAGCTTCGGGGAGTGCAGCCCAAGCCTT 899
QY 794 GGCAGCTTCCAAAGTGTGTGAGCCTGTGGGTGCAAGAAGTCAAGAATTTGAGTTTGGG 853
DB 898 GACAGCTTCCATGTGTGTGTGAGACTGCGAGTGCAGAGAAGTCAAGAATTTGGGTTTGA 839
QY 854 AACCTCCAATCAGATTCAGAGATATATGGAACCCCTGGATGCCAGGCAAGATTGG 913
DB 838 AACCTTCGCCTAGATTAAGAGAGATGTGGGAAATTCCTGGATGCCAGTCAAGAAATTTG 779
QY 914 CTGTAGGGTGGGTCCTCATGAGAACCTCTGCAAGGGTACTACAAAGGCAAAATGTTG 973
DB 778 CTGACGAGACAGGGCCCTCATGGATCTCTGCGACGGCAGTGCAGAAAGGAAATGTTG 719
QY 974 GGTGGAGGCCCCACACAGAGTCCCAAGTGGGCTCCATCTAGTAGAGCTGTGAGAAGAA 1033
DB 718 GGTGACAGACCCCAACACAGTCCCTACTGTGGGCACCACTAGTGGAGCTGTGAGAAG 659
QY 1034 GTCCACCATCTCCAGACTCCAGAGGGTAGATCCACTGACAGCTTGCAGATGTGCGCTG 1093
DB 658 G-----TCCTCCAGACCCCAAGATGGTAGATCCACCCGACAGCTTGCACCGCTGACCTG 606

QY 1094 AAAAAATCCACAGACACTCAGTCCAGCCTGTGAAAGCAGCAGGATGGAGTCTGTACCCCT 1153
DB 605 GAAAGCTGCACACACTCAATCCAGCCCATGAAGACAGCTGAGAGGAGGCTGTACCCCT 546
QY 1154 ACAAAACCGTAGTGGCAGAGCTGACACAGACCGTGGGAATCTACCTCTTTCATTTGTCATG 1213
DB 545 GCAAAGGTACAGGGGCAGAGCTGCCCAAGACCATGGGAACCCCTTCCATCTCGGTG 486
QY 1214 ACCTGGACCTGAGACATGAGTCAAAAGAGATCATTTTGGAGCTTTAAGATTTGACTGCC 1273
DB 485 ACCTGGATGTGAGATGTGGAGTCAAAAGGAGATCATTTTGGAGCTTTAAGATTTGACTGCC 426
QY 1274 CCACCTGGATTTCCGAGCTTATATATGGGCGCCGTACCCCTTTGTTTGGCCCAATTTTCCAT 1333
DB 425 CCACCTGGATTTTGGACTCTCATGGGCGCTGTAGCCCTTTGTTTGGACCAATTTATCCCAT 366
QY 1334 TTGGAATGCGCTATTTATCCCAATGCTGTACTCATTCATTTGAT 1376
DB 365 TTGGAATGCGCTATTTATCCCAATGCTGTACTCCCATTTGAT 323

RESULT 6
AAS03689/c
ID AAS03689 standard; DNA; 9236 BP.
XX
AC AAS03689;
XX
DT 29-AUG-2001 (first entry)
XX
DE Rhesus gene locus: upstream Rhesus box of D-positives.
XX
KW Rhesus box; RHD positive; sequence length polymorphism; SSP: RHD; SMPL:
KW RHCE; Rh negative; blood group typing; blood transfusion; antigen C;
KW haemolytic disease of the newborn; chromosome 1 p34.1-p36; ds.
XX Homo sapiens.
XX WO200132702-A2.
XX
PD 10-MAY-2001.
XX
PF 31-OCT-2000; 2000WO-EP10745.
XX
PR 02-NOV-1999; 99EP-0121686.
PR 31-MAY-2000; 2000EP-0111696.
XX
PA (DRKB-) DRK BLUTSPENDEDIENST BADEN WUERTEMBERG.
XX
PI Flegel WA, Wagner RF;
XX
DR WPI; 2001-291052/30.
XX
PT New nucleic acid molecular structure, useful for detection of common
PT RHD positive haplotypes in D-negative individuals, comprises RHD, SMPI
PT and RHCE genes -
XX
PS Disclosure: Fig 9; 135pp; English.
XX
CC The sequence represents the coding sequence of Rhesus gene locus:
CC upstream Rhesus box of D positives. The Rhesus genes locus
CC comprises the RHD, SMPI and RHCE (all undefined) genes and/or the
CC Rhesus box(es), preferably the hybrid Rhesus box, the upstream Rhesus box
CC and/or the downstream Rhesus box. The RHD and RHCE genes are located at
CC chromosome 1 p34.1-p36. Rhesus box flanks the breakpoint region of the
CC RHD deletion in the common RHD negative haplotypes. The sequence has
CC been used to design primers which are useful for: (1) the specific
CC detection of the common RHD positive haplotypes in D-negative
CC individuals; (2) blood group typing; (3) determining whether a patient
CC can be transfused with RhD negative blood and whether blood is suitable
CC for transfusion to patients who should not be exposed to antigen C; (4)
CC assessing the risk of a RhD negative mother of conceiving or carrying an
CC RhD positive foetus. Anti-D antibodies are useful for treating pregnant
CC RhD positive foetus. Anti-D antibodies are useful for treating pregnant

QY 411 AAGCATTCAGAGGTGACTTGGTGCTGTAAAGGCATTCAGTTTCATAAGGAGGCAGA 470
Db 5993 AAGCATTCAGAGGTGACTTGGTGCTGTAAAGGCATTCAGTTTCATAAGGAGGCAGA 5934
QY 471 GCATAAGAGTTCAGAAAATTTCCACCCCTGCACAAATGTGATAAAAAAGAAAA--CCCATTT 528
Db 5933 GCATAAAGTTCGGAATTTCCAGCCCTGCACAAATGTGATAAAAAAGAAAAATTCACATTT 5874
QY 529 TCTGAGGGAAATTCAGCTGGCTGCAGAAATTTGCATATGTAAATCAGGAGCTGAATGTT 588
Db 5873 TCTGAGGAGAAATTCAGAGTGCCTGCAGAAATTTGCATGAGTAA-CAGGAGCCAAATGCT 5815
QY 589 AATCCTCAAGACAATGGGGAATATTCCTCGACATGTCAAGAGTCTTCACAGAGTCC 648
Db 5814 AATTCCTCAAGACAATGGGGAATATTCCTCGACATGTCAAGAGTCTTCACAGAGTCC 5755
QY 649 ATCAATCACTGGCTGGAGGCT-AGGAGAAATGGTTTTCTGGACAGGCCAGGGTC 707
Db 5754 CTCCCATCAGAGTCCAGAGGTATCAGGAAATATGGTTTTCTGGCCAGGCCGGGGTC 5695
QY 708 CCGTGTCTGTGCAGCTTAGAGACTTGGTGCCCTGTGTCCAGTTA----- 754
Db 5694 CTCATGCTGTGCAGCTTAGGACTTGGTGCCCTGCATCCAGCCACTCCCAACCATGA 5635
QY 755 -----ATTGAGTGTGGCTTCAGAGGGTGCAAGCCCAAGCCTT 793
Db 5634 CTGACGGGAGCAAGGTAGAGCTTGGGCTGTAGCTTCGGGGAGTGCAAGCCCAAGCCTT 5575
QY 794 GCGAGTTCAGAGTGGTGTGAGCTTGGGTGCAAGAGTCAAGATTCAGAGTTTGGG 853
Db 5574 GACAGCTCCATGTGGTGTGAGATGGGAGTGCAAGAGTCAAGATTCAGAGTTTGGG 5515
QY 854 AACCTCAATCAGATTTCAGAGATATATGAAACCCCTGGATGCCAGGCAGAGTTTG 913
Db 5514 AACCTCGCCTAGATTAAAGAGATGTGGGAATGCCCTGGATGCCAGTCAAGATTTC 5455
QY 914 CTGTAGGGTGGGTCTCATGGAGAACTCTGCAAGGGTAGTACAAAAGGAAATGTTG 973
Db 5454 CTGACGAGCAGGGCCCTCATGGAGATCTCTGCCAGGCACTGCAGAGGGAATGTGG 5395
QY 974 GGTGGAGCCCCACACAGAGTCCCAGTGGGCTCCATCTAGTAGAGCTGTGAGAGAA 1033
Db 5394 GGTGAGAGACCCACACAGATCCCTACTGGGGCCACCACTAGTGGAGCTGTGAGAGAG 5335
QY 1034 GTCCACATCTCCAGACTCCAGAGGGTAGATCCACTGACAGCTGCAGAGTGTGCCCTG 1093
Db 5334 G-----TCCTCCAGACCCAGATGGTAGATCCACGACAGCTGACCTGTACCTG 5282
QY 1094 AAAATCCAGACACTCAGTGCAGCCTGTGAAAGCAGCAGGGATGTGACCTT 1153
Db 5281 GAAAGCTGCAGACACTCAATGCCAGCCCATGAAAGCAGCTGAGAGGGAGCTGACCT 5222
QY 1154 ACAAAACCTAGTGGCAGAGCTGACCAAGACCGTGGGAATCTACCTTTCATGTCATG 1213
Db 5221 GAAAGGTACAGGGGACAGCTGCCCAAGACCATGGGAACCCACCCCTTCCATCTCGCTG 5162
QY 1214 ACCTGGACCTGAGACTGAGGTCAAAAGAGATCATTTTGGAGCTTTAAGATTTGACTGCC 1273
Db 5161 ACCTGGATCTGAGATCTGAGTCAAGAGAGATCATTTTGGAGCTTTAAGATTTGACTGCC 5102
QY 1274 CCATGGATTTCGGACTATATATGGGGCCCGTATCCCTTTGTTTGGCCAAATTTTTCAT 1333
Db 5101 CCATGGATTTCGGACTCTCATGGGCTGTAGCCTCTTGTGTTTGTGACCAATTTATCCCAT 5042
QY 1334 TTGGAACTCCGCTATTACCCCAATGCCTGTACCTCCATGTTAT 1376
Db 5041 TTGGAACTGCTGTATTTACCCCAATGCCTGTACCCCAATGTTAT 4999

RESULT 8
AAS03688/c
ID AAS03688 standard; DNA; 9241 BP.
XX

AC AAS03688;
DT 29-AUG-2001 (first entry)
DE Rhesus gene locus: hybrid Rhesus box of RHD negatives.
KW Rhesus box; RHD positive; sequence length polymorphism; SSP; RHD; SMP1;
KW RHCE; Rh negative; blood group typing; blood transfusion; antigen C;
KW haemolytic disease of the newborn; chromosome 1 p34.1-p36; ds.
OS Homo sapiens.
XX WO200132702-A2.
XX 10-MAY-2001.
XX 31-OCT-2000; 2000MO-EPI0745.
XX 02-NOV-1999; 99EP-0121686.
PR 31-MAY-2000; 2000EP-0111696.
XX (DRKB-) DRK BLUTSPENDEDIENST BADEN WUERTEMBERG.
XX Flegel WA, Wagner PF;
PI WPI; 2001-291052/30.
XX New nucleic acid molecular structure, useful for detection of common
PT RHD positive haplotypes in D-negative individuals, comprises RHD, SMP1
PT and RHCE genes -
XX Disclosure; Fig 8; 135pp; English.
XX The sequence represents the coding sequence of Rhesus gene locus:
CC hybrid Rhesus box of RHD negatives. The Rhesus genes locus
CC comprises the RHD, SMP1 and RHCE (all undefined) genes and/or the
CC Rhesus box(es), preferably the hybrid Rhesus box, the upstream Rhesus box
CC and/or the downstream Rhesus box. The RHD and RHCE genes are located at
CC chromosome 1 p34.1-p36. Rhesus box flanks the breakpoint region of the
CC RHD deletion in the common RHD negative haplotypes. The sequence has
CC been used to design primers which are useful for: (1) the specific
CC detection of the common RHD positive haplotypes in D-negative
CC individuals; (2) blood group typing; (3) determining whether a patient
CC can be transfused with Rhd negative blood and whether blood is suitable
CC for transfusion to patients who should not be exposed to antigen C; (4)
CC assessing the risk of a Rhd negative mother of conceiving or carrying an
CC Rhd positive foetus. Anti-D antibodies are useful for treating pregnant
CC women who are Rhesus D negative, where the foetus is not homozygous for
CC the RHD gene to treat or prevent haemolytic disease of the newborn.
XX
SQ Sequence 9241 BP; 2454 A; 2320 C; 2014 G; 2453 T; 0 other;
Query Match 46.5%; Score 639.8; DB 22; Length 9241;
Best Local Similarity 79.1%; Pred. No. 3e-189;
Matches 841; Conservative 0; Mismatches 177; Indels 45; Gaps 5;
QY 351 GAATATTTGGCTAAATCAGGAAGCAAAATTTTGGCTAATAAGGAAGAATTTCTTAAGCAGA 410
Db 6053 GAACITTTGACITTCAGAGAGATCATTTAGAGTATCTGGCAGGAGAAATTTCTTAAGCAGA 5994
QY 411 AAGCATTCAGAGGTGACTTGGTGCTGTAAAGGCATTCAGTTTCATAAGGAGGCAGA 470
Db 5993 AAGCATTCAGAGGTGACTTGGTGCTGTAAAGGCATTCAGTTTCATAAGGAGGCAGA 5934
QY 471 GCATAAGAGTTCAGAAAATTTGCACCCCTGCACAAATGTGATAAAAAAGAAAA--CCCATTT 528
Db 5933 GCATAAAGTTCGGAATTTTCAGCCCTGCACAAATGTGATAAAAAAGAAAAATTCACATTT 5874
QY 529 TCTGAGGGAAATTCAGCTGGCTGCAGAAATTTGCATATGTAAATCAGGAGCTGAATGTT 588
Db 5873 TCTGAGGAGAAATTCAGCTGGCTGCAGAAATTTGCATGAGTAA-CAGGAGCCAAATGCT 5815
QY 589 AATCCTCAAGACAATGGGGAATATATCCTCGACATGTCAAGAGTCTTCACAGAGTCC 648

[illegible]

WO200216399-A2

XX 28-FEB-2002.
 PD 27-AUG-2001; 2001WO-US26663.
 PF 25-AUG-2000; 2000US-227815P.
 PR (GENA-) GENA/ISSANCE PHARM INC.
 PA Anastasio AE, Kazeml A;
 PI WPI; 2002-280907/32.
 DR P-PSDB; AAU78238.
 XX
 XX Novel isolated polynucleotide which is a polymorphic variant of
 PT tachykinin receptor 1 (TACR1) gene useful for expressing TACR1 protein
 PT isoform used in screening drug candidates to treat pain, depression,
 PT vomiting -
 XX
 XX Claim 1; Flg 1; 89pp; English.
 XX The invention relates to an isolated polynucleotide sequence which
 CC comprises a tachykinin receptor 1 (TACR1) isogene (SG) that is any one of
 CC 16 SG as given in specification, where each SG comprises specific regions
 CC of the TACR1 genomic DNA appearing as ABK12169, and is defined by
 CC polymorphisms at positions (P) 3164, 3319, 3906, 4339, 4444, 92915,
 CC 94601, 94821, 94892, 94960. Also included are fragments of the
 CC TACR1 isogenes and TACR1 cDNA, a transgenic non-human animal transformed
 CC with the TACR1 isogene or coding region, haplotyping (or genotyping) the
 CC TACR1 of an individual by determining either the haplotype of one or both
 CC copies of the TACR1 gene, predicting the haplotype pair for the
 CC TACR1 gene of an individual, identifying an association between a trait
 CC and a haplotype pair, an isolated oligonucleotide for detecting the
 CC polymorphisms, a computer system for storing and analysing polymorphism
 CC data and a genome anchology for TACR1 gene. The TACR1 isogene is useful
 CC for studying expression and function of TACR1 and expressing TACR1
 CC protein for use in screening for candidate drugs to treat diseases
 CC related to TACR1 activity. The polymorphism and haplotype data is useful
 CC for validating whether TACR1 is a suitable target for drugs to treat
 CC pain, depression, vomiting, acute inflammatory diarrhoea and opiate
 CC addiction, screening for such drugs and reducing bias in clinical trials
 CC of such drugs. The genotyping method is useful for determining whether an
 CC individual has one of the haplotype pairs. The haplotyping method is
 CC useful for improving efficiency and outcome of several steps in
 CC discovery and development of drugs for treating the diseases. The
 CC haplotyping method is also useful for validating TACR1 as a candidate
 CC target for treating a specific condition or disease predicted to be
 CC associated with TACR1 activity. The method is also useful for screening
 CC compounds to treat a specific condition or disease predicted to be
 CC associated with TACR1 activity. The methods are useful for identifying
 CC an association between susceptibility to a disease, staging of a disease,
 CC or response to a drug. The gene for TACR1 is located on human
 CC chromosome 2. The present sequence is a representative gene sequence
 CC for TACR1.
 XX
 XX Sequence 98690 BP; 29293 A; 20385 C; 20052 G; 28885 T; 75 other;
 SQ
 Query Match 45.8%; Score 631; DB 24; Length 98690;
 Best Local Similarity 80.2%; Pred. No. 7.le-186;
 Matches 820; Conservative 0; Mismatches 160; Indels 43; Gaps 5;
 QY 391 GAAGAAATTTCTAAGCAGCAAGCATTCAGAGGTGACTTGGGTGCTGTAAAGGCATTTC 450
 DB 51906 GAAGAAATTTCTAAGCAGCAAGCATTCAGAGGTGACTTGGGTGCTGTAAAGGCATTTC 51847
 QY 451 AGTTTCATAAGGGAGGAGAGCATTAAGAGTTTCAGAAAAATTTGCACCCCTGACAAATGTGATA 510
 DB 51846 AGTTTCATAAGGGAGGAGAGCATTAAGAGTTTCAGAAAAATTTGTAGCTCTGACTATCGGATA 5187
 QY 511 AAAAAAGAAAAACCCATTCTCTCAGGGGAAATTCAGAGCTGGCTGCAGAAATTTGCATATGT 570
 DB 51786 GAAAAAGAAAACTCATTTCTCTGGGGAGAGAAATTCAGAGCTGGCTGCAGAAATTTGCATAAGT 51727

QY 571 AATGAGGAGCTGAATGTTAAATCCTCAAGACAATGGGAAAAATATCTCCTGGACATGTCTCAG 630
 DB 51726 AGCAGGAGCCCTAATGTTAAATCCTCAAGACAATGGGAAAAATGTTCTCCAGGTGATGTCTCAG 51667
 QY 631 AGGTCTTACACAGCAGTCCATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAAT 687
 DB 51666 AGACCTTCATGCGAGCCCTCCGATCACAGGCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 51607
 QY 688 TGTGGGACAGGCCAGGGTCCCTGCTGT 747
 DB 51606 TGTGGATCTGGCCAGGGTCCCTGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 51547
 QY 748 CCAGTTAATTCAGCTGTGG-----CTTCAGAG 774
 DB 51546 CCAGCCACTCCAGCTGTGGCTGAAAGGGGCCAATGTACAGCTTGGGCTGTGACTTCAGAG 51487
 QY 775 GGTGCAAGCCCAAGCCCTTGGCAGCTTCCAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 834
 DB 51486 GATGCAAGCCCAAGCCCTTGGCAGCTTCCACATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 51427
 QY 835 TCAAGAAATTTAGGTTTGGGAACCTCCAATCAGATTTTCAGAAATATATGTGAAACCCCTGG 894
 DB 51426 TCAAGAAATTTAGATTTGGGAACCTTTTGCTAGATTTAGGAAGACGTATGTGAAAGTACCTGG 51367
 QY 895 ATGCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 954
 DB 51366 ATGCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 51307
 QY 955 GTACAAAAAGGAAATTTGGTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1014
 DB 51306 GTGTGGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 51247
 QY 1015 AGTAGAGCTGTGAGAAGAGTCCACCATCTCCAGACTCCAGAGGGGTAGATCCACTGAC 1074
 DB 51246 AGTTGAGCTGTGAGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 51187
 QY 1075 AGCTTTCAGCATGTGCTCT-----GAAAAATCCACAGACACTCAGTCCGAGCCCTGTGAAAGCAGC 1133
 DB 51186 AGCTTTCAGCATGTGCTCTGGAAGAGGCTGTAGACACTCAACACAGCCCTGTGAAAGTAGC 51127
 QY 1134 AGGGATGAGCTGTACCCCTCAAAACCGTAGTGGGAGAGCTGACCAAGACCGTGGGAAT 1193
 DB 51126 TGGGAG 51067
 QY 1194 CTACCTCTGCAATTTGCTGACCTGAGACCTGAGACATGAGATGAGATGAGATGAGATGAGATGAGAT 1253
 DB 51066 CTACCTCTGCAATTTGCTGACCTGAGACCTGAGACCTGAGACCTGAGACCTGAGACCTGAGACCTGAG 51007
 QY 1254 AGCTTTAAGATTTGACTGCCCCACTGGATTTTCGGACTTATATGGGCGCGGTACCCCTTTG 1313
 DB 51006 AACTTTAAAAATTTGACTG-CCCATTTGATTTTCAGACTTGCATGAGCCCTGTA-----TTG 50953
 QY 1314 TTTTGGCCAAATTTTTCATTTGGAACTGCGCTGATTTTACCCCAATGCTGTACCTGCATTTG 1373
 DB 50952 TTTTGGCCAAATTTTTCATTTGGAACTGCGCTGATTTTACCCCAATGCTGTACCTGCATTTG 50893
 QY 1374 TAT 1376
 DB 50892 TGT 50890
 RESULT 10
 AAL03635
 ID AAL03635 standard; DNA; 6063 BP.
 XX
 AC AAL03635;
 XX
 DT 21-NOV-2001 (first entry)
 XX
 DE Human reproductive system related antigen DNA SEQ ID NO: 6323.
 XX
 KW Human reproductive system related antigen; reproductive system disorder;
 KW cancer; gene therapy; ds.

XX OS Homo sapiens.
XX XX WO200155320-A2.
XX XX 02-AUG-2001.
XX PF 17-JAN-2001; 2001WO-US01339.
XX PR 31-JAN-2000; 2000US-0179065.
PR PR 04-FEB-2000; 2000US-0180628.
PR PR 24-FEB-2000; 2000US-0184664.
PR PR 02-MAR-2000; 2000US-0186350.
PR PR 16-MAR-2000; 2000US-0189874.
PR PR 17-MAR-2000; 2000US-0190076.
PR PR 18-APR-2000; 2000US-0198123.
PR PR 19-MAY-2000; 2000US-0205515.
PR PR 07-JUN-2000; 2000US-0209467.
PR PR 28-JUN-2000; 2000US-0214886.
PR PR 30-JUN-2000; 2000US-0215135.
PR PR 07-JUL-2000; 2000US-0216647.
PR PR 07-JUL-2000; 2000US-0216880.
PR PR 11-JUL-2000; 2000US-0217487.
PR PR 11-JUL-2000; 2000US-0217496.
PR PR 14-JUL-2000; 2000US-0218290.
PR PR 26-JUL-2000; 2000US-0220963.
PR PR 26-JUL-2000; 2000US-0220964.
PR PR 14-AUG-2000; 2000US-0224518.
PR PR 14-AUG-2000; 2000US-0224519.
PR PR 14-AUG-2000; 2000US-0225213.
PR PR 14-AUG-2000; 2000US-0225214.
PR PR 14-AUG-2000; 2000US-0225266.
PR PR 14-AUG-2000; 2000US-0225267.
PR PR 14-AUG-2000; 2000US-0225268.
PR PR 14-AUG-2000; 2000US-0225270.
PR PR 14-AUG-2000; 2000US-0225477.
PR PR 14-AUG-2000; 2000US-0225757.
PR PR 14-AUG-2000; 2000US-0225758.
PR PR 14-AUG-2000; 2000US-0225759.
PR PR 18-AUG-2000; 2000US-0226279.
PR PR 22-AUG-2000; 2000US-0226681.
PR PR 22-AUG-2000; 2000US-0226688.
PR PR 22-AUG-2000; 2000US-0227182.
PR PR 23-AUG-2000; 2000US-0227009.
PR PR 30-AUG-2000; 2000US-0228924.
PR PR 01-SEP-2000; 2000US-0229287.
PR PR 01-SEP-2000; 2000US-0229343.
PR PR 01-SEP-2000; 2000US-0229344.
PR PR 01-SEP-2000; 2000US-0229345.
PR PR 05-SEP-2000; 2000US-0229509.
PR PR 05-SEP-2000; 2000US-0229513.
PR PR 06-SEP-2000; 2000US-0230437.
PR PR 06-SEP-2000; 2000US-0230438.
PR PR 08-SEP-2000; 2000US-0231242.
PR PR 08-SEP-2000; 2000US-0231243.
PR PR 08-SEP-2000; 2000US-0231244.
PR PR 08-SEP-2000; 2000US-0231413.
PR PR 08-SEP-2000; 2000US-0231414.
PR PR 08-SEP-2000; 2000US-0232080.
PR PR 08-SEP-2000; 2000US-0232081.
PR PR 12-SEP-2000; 2000US-0231968.
PR PR 14-SEP-2000; 2000US-0232397.
PR PR 14-SEP-2000; 2000US-0232398.
PR PR 14-SEP-2000; 2000US-0232399.
PR PR 14-SEP-2000; 2000US-0232400.
PR PR 14-SEP-2000; 2000US-0232401.
PR PR 14-SEP-2000; 2000US-0233063.
PR PR 14-SEP-2000; 2000US-0233064.
PR PR 21-SEP-2000; 2000US-0233065.
PR PR 21-SEP-2000; 2000US-0234223.
PR PR 21-SEP-2000; 2000US-0234274.
PR PR 25-SEP-2000; 2000US-0234997.
PR PR 25-SEP-2000; 2000US-0234998.
PR PR 25-SEP-2000; 2000US-0234999.
PR PR 26-SEP-2000; 2000US-0235484.
PR PR 27-SEP-2000; 2000US-0235834.
PR PR 27-SEP-2000; 2000US-0235836.
PR PR 29-SEP-2000; 2000US-0236327.
PR PR 29-SEP-2000; 2000US-0236367.
PR PR 29-SEP-2000; 2000US-0236368.
PR PR 29-SEP-2000; 2000US-0236369.
PR PR 29-SEP-2000; 2000US-0236370.
PR PR 02-OCT-2000; 2000US-0236802.
PR PR 02-OCT-2000; 2000US-0237037.
PR PR 02-OCT-2000; 2000US-0237038.
PR PR 02-OCT-2000; 2000US-0237039.
PR PR 02-OCT-2000; 2000US-0237040.
PR PR 13-OCT-2000; 2000US-0239935.
PR PR 13-OCT-2000; 2000US-0239937.
PR PR 20-OCT-2000; 2000US-0240960.
PR PR 20-OCT-2000; 2000US-0241221.
PR PR 20-OCT-2000; 2000US-0241785.
PR PR 20-OCT-2000; 2000US-0241786.
PR PR 20-OCT-2000; 2000US-0241787.
PR PR 20-OCT-2000; 2000US-0241808.
PR PR 20-OCT-2000; 2000US-0241809.
PR PR 20-OCT-2000; 2000US-0241826.
PR PR 01-NOV-2000; 2000US-0244617.
PR PR 08-NOV-2000; 2000US-0246474.
PR PR 08-NOV-2000; 2000US-0246475.
PR PR 08-NOV-2000; 2000US-0246476.
PR PR 08-NOV-2000; 2000US-0246477.
PR PR 08-NOV-2000; 2000US-0246478.
PR PR 08-NOV-2000; 2000US-0246523.
PR PR 08-NOV-2000; 2000US-0246524.
PR PR 08-NOV-2000; 2000US-0246525.
PR PR 08-NOV-2000; 2000US-0246526.
PR PR 08-NOV-2000; 2000US-0246527.
PR PR 08-NOV-2000; 2000US-0246528.
PR PR 08-NOV-2000; 2000US-0246532.
PR PR 08-NOV-2000; 2000US-0246609.
PR PR 08-NOV-2000; 2000US-0246610.
PR PR 08-NOV-2000; 2000US-0246611.
PR PR 08-NOV-2000; 2000US-0246613.
PR PR 17-NOV-2000; 2000US-0249207.
PR PR 17-NOV-2000; 2000US-0249208.
PR PR 17-NOV-2000; 2000US-0249209.
PR PR 17-NOV-2000; 2000US-0249210.
PR PR 17-NOV-2000; 2000US-0249211.
PR PR 17-NOV-2000; 2000US-0249212.
PR PR 17-NOV-2000; 2000US-0249213.
PR PR 17-NOV-2000; 2000US-0249214.
PR PR 17-NOV-2000; 2000US-0249215.
PR PR 17-NOV-2000; 2000US-0249216.
PR PR 17-NOV-2000; 2000US-0249217.
PR PR 17-NOV-2000; 2000US-0249218.
PR PR 17-NOV-2000; 2000US-0249219.
PR PR 17-NOV-2000; 2000US-0249244.
PR PR 17-NOV-2000; 2000US-0249245.
PR PR 17-NOV-2000; 2000US-0249264.
PR PR 17-NOV-2000; 2000US-0249265.
PR PR 17-NOV-2000; 2000US-0249297.
PR PR 17-NOV-2000; 2000US-0249299.
PR PR 17-NOV-2000; 2000US-0249300.
PR PR 01-DEC-2000; 2000US-0250160.
PR PR 01-DEC-2000; 2000US-0250391.
PR PR 05-DEC-2000; 2000US-0251030.
PR PR 05-DEC-2000; 2000US-0251988.
PR PR 05-DEC-2000; 2000US-0256719.
PR PR 06-DEC-2000; 2000US-0251479.
PR PR 08-DEC-2000; 2000US-0251856.
PR PR 08-DEC-2000; 2000US-0251868.
PR PR 08-DEC-2000; 2000US-0251869.
PR PR 08-DEC-2000; 2000US-0251989.
PR PR 08-DEC-2000; 2000US-0251990.
PR PR 11-DEC-2000; 2000US-0254097.
PR PR 05-JAN-2001; 2001US-0259678.
XX XX


```
FT /tag= b
FT /number= 3
FT /cons_splice= (5'site:YES,3'site:NO)
FT replace(64331,G)
FT /tag= a
FT /standard_name= "Single nucleotide polymorphism"
FT 166228..166322
FT /tag= b
FT /number= 4
FT 166323..168001
FT /tag= b
FT /number= 4
FT replace(167950,G)
FT /tag= a
FT /standard_name= "Single nucleotide polymorphism"
FT replace(167989,G)
FT /tag= a
FT /standard_name= "Single nucleotide polymorphism"
FT 168002..168120
FT /tag= b
FT /number= 5
FT replace(168054,G)
FT /tag= a
FT /standard_name= "Single nucleotide polymorphism"
FT 168121..169542
FT /tag= b
FT /number= 5
FT /cons_splice= (5'site:YES,3'site:NO)
FT 169543..169825
FT /tag= b
FT /number= 6
FT replace(169812,G)
FT /tag= a
FT /standard_name= "Single nucleotide polymorphism"
FT replace(169823,G)
FT /tag= a
FT /standard_name= "Single nucleotide polymorphism"
FT 169826..169866
FT /tag= b
FT /number= 6
FT /cons_splice= (5'site:YES,3'site:NO)
FT 169867..170678
FT /tag= b
FT /number= 7
FT replace(170035,A)
FT /tag= a
FT /standard_name= "Single nucleotide polymorphism"
FT replace(170068,T)
FT /tag= a
FT /standard_name= "Single nucleotide polymorphism"
FT replace(170256,C)
FT /tag= a
FT /standard_name= "Single nucleotide polymorphism"
FT replace(170368,G)
FT /tag= a
FT /standard_name= "Single nucleotide polymorphism"
FT replace(170487,C)
FT /tag= a
FT /standard_name= "Single nucleotide polymorphism"
FT 170679..204911
FT /tag= b
FT /number= 7
FT 204912..205102
FT /tag= b
FT /number= 8
FT 205103..242969
FT /tag= b
FT /number= 8
FT 242970..243086
FT /tag= b
FT /number= 9
FT replace(243055,T)
FT /tag= a
```

```
FT intron
FT /standard_name= "Single nucleotide polymorphism"
FT 243087..306167
FT /tag= b
FT /number= 9
FT replace(243187,C)
FT /tag= a
FT /standard_name= "Single nucleotide polymorphism"
FT 306168..306503
FT /tag= b
FT /number= 10
FT replace(306292,A)
FT /tag= a
FT /standard_name= "Single nucleotide polymorphism"
FT replace(306382,G)
FT /tag= a
FT /standard_name= "Single nucleotide polymorphism"
FT 306504..373639
FT /tag= b
FT /number= 10
FT 373640..373778
FT /tag= b
FT /number= 11
FT 373779..422963
FT /tag= b
FT /number= 11
FT 422964..423097
FT /tag= b
FT /number= 12
FT replace(423067,C)
FT /tag= a
FT /standard_name= "Single nucleotide polymorphism"
FT 423098..456353
FT /tag= b
FT /number= 12
FT replace(423149,G)
FT /tag= a
FT /standard_name= "Single nucleotide polymorphism"
FT replace(423163,G)
FT /tag= a
FT /standard_name= "Single nucleotide polymorphism"
FT replace(423220,A)
FT /tag= a
FT /standard_name= "Single nucleotide polymorphism"
FT replace(423232,G)
FT /tag= a
FT /standard_name= "Single nucleotide polymorphism"
FT replace(423258,G)
FT /tag= a
FT /standard_name= "Single nucleotide polymorphism"
FT 456354..456537
FT /tag= b
FT /number= 13
FT 456538..460700
FT /tag= b
FT /number= 13
FT replace(459706,C)
FT /tag= a
FT /standard_name= "Single nucleotide polymorphism"
FT replace(459832,A)
FT /tag= a
FT /standard_name= "Single nucleotide polymorphism"
FT replace(459913,G)
FT /tag= a
FT /standard_name= "Single nucleotide polymorphism"
FT replace(460024,G)
FT /tag= a
FT /standard_name= "Single nucleotide polymorphism"
FT 460056,T)
FT /tag= a
FT /standard_name= "Single nucleotide polymorphism"
FT replace(460159,G)
FT /tag= a
FT /standard_name= "Single nucleotide polymorphism"
```


SQ	Sequence	465237 BP;	133988 A;	89578 C;	93946 G;	147721 T;	4 other:
	Query Match	44.9%;	Score	618.2;	DB	24;	Length 465237;
	Best Local Similarity	78.2%;	Pred.	No.	1.9e-181;		
	Matches	847;	Conservative	0;	Mismatches	193;	Indels 43; Gaps
QY	334	CTGCATCATTTGCAAGGTGAATAATTTGGCTAATGAGGAACAAATTTTGCTCATTAAAGAA	393				
Db	438730	CTGCTGTAGAGATTGTGGAAATTTGAACATTGAGAGAAATGATTTGGGTACCTGGTAAAA	438789				
QY	394	GAAATTTCTAAGCAGCAACCATTCACAGAGTGACTTGGTGCTGTAAAGCATTCAGT	453				
Db	438790	GAAATTTCTAAGCAGCAACCATTCACAGAGTGACTTGGTGCTGTAAAGCATTCAGT	438849				
QY	454	TTCATAGGAGGACGACGACATAGAGTTCAGAAAATTTGCACCCTCACAAATGTGATAAAA	513				
Db	438850	TT--TAAAGAGAACAAGCATAAAGTTTCAGAAAATTTGCAGCCCTCATGATCGAGTAGGA	438907				
QY	514	AAGAAAACCCA-TTTTTGAGGGGAAATTCAGAGTGGCTGCAGAAAATTTGCATATGTAA	572				
Db	438908	AAGAAAACCCAATTTTTTTCAGGAGAAATTCACAGTGGCTGCAGAAAATTTGCATTAAGTAA	438967				
QY	573	TGAGGAGCTCAATGTTAATCCTCAACACAAATGGGAAATATCTCTGCACATGTCAGAG	632				
Db	438968	CAAGGAGCAAATGTTAATCCCCAACACAAATGGGAAATATCTCTGCAGAGCATGTCATAG	439027				
QY	633	GTCTTCACAGCAGTCCATCAATCACTGGCCTGGAGCCCTAGGAG-----AAATGGTTTT	688				
Db	439028	GTCTTCATGCAGCCCTCCATCAAGACCCGGAAGCCCTAGGAGGAAAANAACAGTTTTT	439087				
QY	689	GTGGGACAGGCCAGGGTCCCTGTCTGTGTGCAGCCTAGAGCACTTGGTGCCCTGTGTCC	748				
Db	439088	GTGGGCCAGTCCCAGGGTCCCATGCTGTGTGCAGCCTAGGAACCTTGGTGCCCTGCATCT	439147				
QY	749	CAGTTAATTCAGCTGT-----GGCTTCAGAGGGTGCAGGCC	784				
Db	439148	CAGTGTCTCAGCATATTTGCTTAAAGGGGGTGAAGTACCACGGTTTTTCAGAGGTTCGAAGCC	439207				
QY	785	CCAAGCCTTGGCAGCTTCCAAGTGGTGTGAGCCTGTGGGTGCAAAAGAGTCAAGAATTG	844				
Db	439208	CCAACCTTGGCAGCTTTCATGTGGTGTGAGCCTGTGTACACAGAAAGTTAAGAAATTG	439267				
QY	845	AGTTTGGGAACCTCCAATCAGATTTCAGAAGATATATGAAACCCCTGGATGCCCAGGC	904				
Db	439268	AGTTTGGGAACCTCCACTTATATTTTCAGAAGATACGTGGAATGCTTGGATACCCAGGC	439327				
QY	905	AGAAATTTGCTGTAGGGTGGGTCCCTCATGNGAACCCTCTGCAAGGTTAGTACAAAAGG	964				
Db	439328	AAACATTTGCTGACAGAGTGGGGCCCTCATGGAGGCCCTCTGCTAGGGCAATGAGGAAGG	439387				
QY	965	GAAATTTGGTGGGAGGCCCCACACAGAGTCCCACAGTGGGGCTCCATCTAGTAGAGCTG	1024				
Db	439388	GAAATTTGGTGGGAGGCCCCACACAGAGTCCCACAGTGGGGCTCCATCTAGTAGAGCTG	439447				
QY	1025	TGAGAAAGAGTCCACCATCCTCCAGACTCCAGAGGTTAGATCCACTGACAGCTTGCAGC	1084				
Db	439448	TAAGAGSAGGACCACTGTCTCTCCAGACCACAGAA TAGTAGATCCACTGACAGCTTGCA	439507				
QY	1085	ATGTGCTTGAAAATTCACACAGACATCTAGTGCCAGCCTGTGAAGCA-----GCA	1134				
Db	439508	ATGTGCTTGAAAAGCCACAGACATCTAAGCCAGCCTGTGAAGCAGTTCAGGGTTGGAG	439567				
QY	1135	GGATGGAGTCTACCTTACAAAACCTGATGGCAGAGCTGACCAAGACCCGTGGGAATC	1194				
Db	439568	GTGGTGGTGGCTATACCTTATAAGGCCACAGGGGACAGCTGCCCAAGACTATGGGAACC	439627				
QY	1195	TACCTCTTGCATTTGTCATGACCTGGACGTGAGACATGGAGTCAAAAGAGATCATTTTGA	1254				
Db	439628	TACCTCTTGCATGAGCATGACCTGGATGTGAGACATTCAGTCAAGAGAGT-ATTTTGA	439686				
QY	1255	GCTTTAAGATTTGACTGCCCACTGGATTTCCGACHTTATATGGGGCCCGT-ACCCCTTTG	1313				
Db	439687	GCTTTAAGATTTGACTGCCCACTGGATTTTACAGATTTGTTGGGCTCTGTAACCCCTTGG	439746				

QY	1314	TTTTGGCCCAATTTTTCACATTTGGAACTGCGGTATTTATTTACCAATGCTGTACCTTCATTTG	1373
Db	439747	TTTTTGGCCCAATTTTCCTCCCAATTTGGAGCTGCTGTATTTATTTACCAATGCTTAAACCGCATTTG	439806
QY	1374	TAT 1376	
Db	439807		
Db	439807	TAT 439809	
RESULT 13			
AAD05134			
ID	AAD05134	standard; cDNA; 3203 BP.	
XX	AC	AAD05134;	
XX	DT	17-JUL-2001 (first entry)	
XX	DE	Human secreted protein-encoding gene 14 CDNA clone H15BF60, SEQ ID NO:24	
XX	KW	Human; secreted protein; proliferative disorder; cancer; tumour;	
XX	KW	fetotal abnormality; developmental abnormality; haematopoietic disorder;	
XX	KW	immune system disorder; AIDS; autoimmune disease; rheumatoid arthritis;	
XX	KW	inflammation; allergy; neurological disorder; Alzheimer's disease;	
XX	KW	Parkinson's disease; cognitive disorder; schizophrenia; asthma;	
XX	KW	skin disorder; psoriasis; sepsis; diabetes; atherosclerosis;	
XX	KW	cardiovascular disorder; angioleptic disorder; kidney disorder;	
XX	KW	gastrointestinal disorder; pregnancy-related disorder;	
XX	KW	endocrine disorder; infection; wound healing; vulnery;	
XX	KW	cell culture; chemotaxis; food additive; gene therapy;	
XX	KW	binding partner identification; ss.	
OS	Homo sapiens.		
XX	Key	Location/Qualifiers	
XX	FT	153..356	
XX	FT	/*tag= a	
XX	FT	/product= "Human secreted protein"	
XX	FT	153..215	
XX	FT	/*tag= b	
XX	FT	mat_peptide	
XX	FT	216..353	
XX	FT	/*tag= c	
XX	FT	/product= "Human mature secreted protein"	
XX	XX		
PN	WO200134769-A2.		
XX	XX		
XX	PD	17-MAY-2001.	
XX	XX		
XX	PF	01-NOV-2000; 2000WO-US30040.	
XX	XX		
PR	05-NOV-1999; 99US-0163580.		
PR	30-JUN-2000; 2000US-0215130.		
XX	XX		
PA	(HUMA-) HUMAN GENOME SCI INC.		
XX	XX		
PI	Ruben SM, Komatsoulis GA, Wei P, Baker KP, Fiscella M;		
XX	XX		
XX	WPI; 2001-308781/32.		
DR	P-PSDB; AAE01245.		
XX	XX		
PT	New isolated nucleic acid molecule encoding a human secreted protein is		
PT	used in preventing, treating or ameliorating a medical condition -		
XX	XX		
PS	Claim 1; Page 408-409; 519pp; English.		
XX	XX		
CC	AAD05121-AAD05203 represent cDNAs corresponding to 24 human secreted		
CC	protein genes, and AAE01232-AAE01311 represent the proteins they encode.		
CC	AAE01312-AAE01340 represent human secreted protein variants or fragments		
CC	The secreted proteins and their genes are useful for preventing,		
CC	treating or ameliorating medical conditions, e.g., by protein or gene		
CC	therapy. Pathological conditions can be diagnosed by determining the		
CC	amount of the new protein in a sample or by determining the presence of		
CC	mutations in the new genes. Specific uses are described for each of the		

24 genes, based on the tissues in which they are most highly expressed, and include developing products for the diagnosis or treatment of proliferative disorders, cancer, tumours, foetal and developmental abnormalities, haematopoietic disorders, diseases of the immune system, AIDS, autoimmune diseases (e.g., rheumatoid arthritis), inflammation, allergies, neurological disorders (e.g., Alzheimer's disease, Parkinson's disease), cognitive disorders, schizophrenia, asthma, skin disorders (e.g., psoriasis), sepsis, diabetes, atherosclerosis, cardiovascular disorders, angioinetic disorders, kidney disorders, gastrointestinal disorders, pregnancy-related disorders, endocrine disorders, and infections. The proteins can also be used to aid wound healing and epithelial cell proliferation, to prevent skin aging due to sunburn, to maintain organs before transplantation, for supporting cell culture of primary tissues, to regenerate tissues, to identify their cognate ligands or binding partners, and in chemotaxis, and can be used as a food additive or preservative to modify storage properties. CC Antibodies specific for a protein of the invention can be used in CC alleviating symptoms associated with the disorders mentioned above, and CC in diagnostic immunoassays e.g., radioimmunoassay or enzyme linked CC immunosorbent assay (ELISA). The present sequence represents a human CC secreted protein-encoding cDNA of the invention.

XX
SQ Sequence 3203 BP; 868 A; 650 C; 740 G; 943 T; 2 other;

Query Match 44.8%; Score 617.4; DB 22; Length 3203;
Best Local Similarity 78.2%; Pred. No. 1.7e-182;
Matches 832; Conservative 0; Mismatches 192; Indels 40; Gaps 6;

QY 351 GAATATTGGCTAATCAGGAGCAAAATTTGGCTAATAGGAGAAATTTCTAAGCAGCA 410
DB 1599 GAACCTTGATCTTGAGAGAGATGATTTAGGGTATCTGGCTGAAGAAATTTCTAAGCAGCA 1658
QY 411 AAGCATTTCAAGAGGTGACCTGGGTGCTGTAAAGGACATTCAGTTTCATAGAGGAGCAGA 470
DB 1659 AAGCATTCAGAGAGTGATTTGGGTGCTGTAAAGGACATTCATTTTAAAGGGA--AACA 1716
QY 471 GCATAAGAGTTCAAGAAATTTGCACCTGCACATGTGATAAAAGAAACCCATTTC 530
DB 1717 GCATAAAATTTGGAAATTTGACGAGCTGTATGATGAGCAGAGAGAGAAACCTGTTTT 1776
QY 531 TGAGGGAAATTCAGCTGGCTGCAGAAATTTGCATATGTAATGAGGAGCTGAATGTTAA 590
DB 1777 TGAGGAGAAATTCAGCTGGCTGCAGAAATTTGCATATGTAATGAGGAGCTGAATGTTAA 1836
QY 591 TCCTCAAGCAATGGGAAATATCTCCTGGACATCTCAGAGTCTTCACAGCAGTCCAT 650
DB 1837 TCCTCAAGCAATGGGAAATATCTCCTGGAGCATCTCAGAGTCTTCACAGTCTTCATGGCAGCCCT 1896
QY 651 CAATCACTGGCTGGAGGCTTAGGAG---AAAATGGTTTGTGGACAGGCCAGGGTC 707
DB 1897 CCCATCACAGACCTCAAGACTAGGAGAAATAATGGTTTGTGGCCAGGCCAGGGTC 1956
QY 708 CTTGTGCTGTGCACGCTAGAGACTTGTGTCCTGTGCTCCAGTTAATTCAG----- 760
DB 1957 CCATGCTGTGTCAGCCTAGGAGACTTGTGTCCTGCATCCCAGCTGCTCCAGCCATTCG 2016
QY 761 -----CTGTGGCTTCAGAGGGTCAAGGCCCAAGCCAGCTGG 795
DB 2017 TAAAGGGCCGAGGTACAGCTCTGCCCATGGTTTCAGAGGGTCAAGCCCAACCTTGG 2076
QY 796 CAGCTTCCAGTGGTGTGAGCCTGTGGGTGCAAGAGTCAAGAAATGAGGTTTGGGAA 855
DB 2077 CAGCTTCCATGTTGTTGAGCCTGCAGGTGTCATGGAAGTCAAGAAATGAGGTTTGGGAA 2136
QY 856 CTTCCAATCAGATTCAGAGATATATGNAACCCCTGATGCCAGGCAAGATTGCT 915
DB 2137 CTTCCAATCAGATTCAGAGATATATGGAATCACCTGATGCTCAGGCAAAAGTTGCT 2196
QY 916 GTAGGGGTGGGTCCTCATGGAGAACTCTGCAAGGCTAGTACAAAAGGGAATGTTGGG 975
DB 2197 CAGGGGCGAGGCTTCATGGAGAACTCTGTAGGGCAGT-CTGAAGGGAAATGTGGG 2255
QY 976 TGGAGGCCCCCACAGAGTCCCACTGGGGGTCCATCTAGTAGAGCTGTGAGAGAAAGT 1035

DB 2256 TTGGAGCCCCCACACAGAAATCCCTACTTGGGGCACCACTAGTAGGAGCTGTSCGAAGAAGG 2315
QY 1036 CCACCATCTCTCCAGACTCCCAAGAGGTAGATCCACTGACAGCTTGACAGCATGTGCCTGAA 1095
DB 2316 CCACCGTCTCTCAGACCCCAAGATGTTAGATCCACGACAGCTGTACCGTGACCTGGA 2375
QY 1096 AATCCACAGACACTCAGTCCAGCTCCAGCTGTGAAAGCAG-CAGGGATGGAGTCTGTACCCCTA 1154
DB 2376 AAGGCCACAGACCTTAAACGCCCGCTGAAAGCAGCAGGAGTGGGGACTATACCCCTG 2435
QY 1155 CAAACCGTAGTGGCAGAGCTGACCAAGACCGTGGGAATCTACTCTTGATTTGTATGA 1214
DB 2436 TGAAGCCACAGGGGCGAGAGCTGCCAAGACTAAGGGAACCTACCTCTTGATCATTTGTA 2495
QY 1215 CTTGAGCTGAGACATGAGTCAAAAGAGATCATTTTGGAGCTTTAAGATTTGACTGCC 1274
DB 2496 CTTGATGTGAGACATGAGTCAAGAGATCATTTTGGAAACGNTATAATTTGACTGCC 2555
QY 1275 CACTGGATTTCCGACTTATATGGGGCCCGT-ACCCCTTTGTTTGGCCAAATTTTCCAT 1333
DB 2556 CGCTGGATTTTGACITGTCATGGGGCTTGTACGCTTTGTTGGGCGAGTTTCTCCCAT 2615
QY 1334 TTGGAAGTCCGCTATTTACCCAATGCTGTACCTCATCTGTATG 1377
DB 2616 TTGGAATGGCTGTATTTACCCATTAACCTGTATCCCATTTGTATG 2659

RESULT 14
AAL03634
ID AAL03634 standard; DNA; 6063 BP.
XX
AC AAL03634;
XX
DT 21-NOV-2001 (first entry)
XX
DE Human reproductive system related antigen DNA SEQ ID NO: 6322.
XX
KW Human; reproductive system related antigen; reproductive system disorder;
KW cancer; gene therapy; ds.
XX
OS Homo sapiens.
XX
PN W0200155320-A2.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01339.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.


```
Db 3559 GCATAAAATTTGGAAAAATTTGAGCCAGTTGATGACGAGAGAGAGAAACCTGTTTTT 3618
QY 531 TGAGGGGAATCAAGCTGGCTGCAGAAATTTGCATATGTAATGAGGAGCTGAATCTAA 590
Db 3619 TGAGGAGAATTCAGCTGGCTGCAGAAATTTGCATAAGTTACAGGAGCTGAATGTTAA 3678
QY 591 TCTCTAAGACAATGGGGAAAAATATCTCTGSGAGATGTCAGAGGTCCTTCACAGACGTCCAT 650
Db 3679 TCTCCAAGACAATGGGGAAAAATGCTCCAGGGCATGTCACAGGTCCTTCATGCGAGCCCT 3738
QY 651 CAAATCACTGGCCCTGGAGCCCTAGAG---AAAATGGTTTTTGTGGACAGGGCCAGGGTC 707
Db 3739 CCATCACAGACCTGAAGACTAGGAGGAAAAATGGTTTTTGTGGCCAGGGCCAGGGTC 3798
QY 708 CTTGTGCTGTGTCAGCCCTAGAGACTTGGTGCCCTGTGTCCTGAGTTAAATTCAG----- 760
Db 3799 CCCATGCTGTGTCAGCCCTAGGAGCTTGGTGCCCTGCATCCAGCTGCTCCAGCCATTGC 3858
QY 761 -----CTGTGGCTTCAGAGGGTGCAGGCCCAAGCCCAAGCCTTGG 795
Db 3859 TAAAGGGCGAGGTACAGCTCTGCCCGTGTTCAGAGGGTGCAGGCCCAAGCCTTGG 3918
QY 796 CAGCTTCCAAGTGGTGTTCAGCCCTGTGGTGCAAGAAGTCAAGAAATTCAGGTTTGGGAA 855
Db 3919 CAGCTTCCAAGTGGTGTTCAGCCCTGCAGGTGCATGGAAGTCAAGAAATTCAGGTTTGGGAA 3978
QY 856 CTCCCAATCAGATTTTCAGAAATATATGAAACCCCTGATGCCAGGCGAGAAAGTTTGGT 915
Db 3979 CCTCCATCTAGATTTTCAGAAAGATGATGGAATCACCCTGATGCTCAGGCAAAAGTTTGGT 4038
QY 916 GTAGGGGTGGGTCTTCATGGAGAACCTCTGCAAGGGTAGTACAAAGGGAATTTGGG 975
Db 4039 CAGGGGCGAGAGCTTCATGAGAACCTCTGCTAGGGCAGT-GTGAAGGGAATTTGGGG 4097
QY 976 TGGGAGCCCCACACAGAGTCCCCAGTGGGGTCCATCTAGTAGAGCTGTGAGAGAAGT 1035
Db 4098 TTGGAGCCCCACACAGAAATCCCTACTGGGGCACCCACTAGTGGAGCTGTGCAAGAGG 4157
QY 1036 CCACATCTCCAGACTCCAGAGGGTAGATCCACTGCAGCTTGCAGCATGTGCTGAA 1095
Db 4158 CCACCTCTCTCAGACCCAGAAATGATAGATCCACGAGAGCTGTACCGTGCACCTGA 4217
QY 1096 AAATCCACAGACACTAGTGGCAGCCTGTGAAAGCAG-CAGGGATGGAGTCTGTACCCCTA 1154
Db 4218 AAGGCCACAGACCTAAAGCCAGCCGCTGAAAGCAGCCAGGAGTGGGGACTATACCCCTG 4277
QY 1155 CAAACCGTAGTGGCAGAGCTGACCAAGACCGTGGGAATCTACCTCTTGCATGTGATGA 1214
Db 4278 TGAAGCCAGGGGCGAGAGCTGCCAAGACTAAGGGAACCTACCTCTTGCATCATTTGA 4337
QY 1215 CTGGACGTGAGACATGGAGTCAAAAGAGATCATTTTGGAGCTTTAAAGATTTGACTGCC 1274
Db 4338 CTGGATGTGAGACATGGAGTCAGAGGAGATCATTTTGGAAAGCTTATATTTGACTGCT 4397
QY 1275 CACTGGATTTGGACATATATATGGGGCCCT-ACCCCTTGTGTTTGGCCAAATTTTCCAT 1333
Db 4398 CGCTGGATTTGGACATGTCATGGGGCTTGAACGCTTGTGTGGGCACTTTCTCCCAT 4457
QY 1334 TTGGAACTGCGGTATTTACCAATGCTGTACCTCCATTTGATG 1377
Db 4458 TTGGAACTGCTGTATTTACCCATTTACCTGTATCCCATTTGATG 4501

RESULT 15
AAH13678
ID AAH13678 standard; cDNA; 1736 BP.
XX
AC AAH13678;
XX
DT 26-JUN-2001 (first entry)
XX
DE Human cDNA sequence SEQ ID NO:10539.
XX
```

Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.

XX Homo sapiens.

PN EP1074617-A2.

XX 07-FEB-2001.

XX 28-JUL-2000; 2000EP-0116126.

XX 29-JUL-1999; 95JP-0248036.

PR 27-AUG-1999; 95JP-0300253.

PR 11-JAN-2000; 2000JP-0118776.

PR 02-MAY-2000; 2000JP-0183767.

PR 09-JUN-2000; 2000JP-0241899.

XX (HELI-) HELIX RES INST.

XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;

PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;

XX WPI; 2001-318749/34.

DR Claim 8: SEQ ID 10539; 2537pp + CD ROM; English.

XX The present invention describes primer sets for synthesizing 5602

CC full-length cDNAs defined in the specification. Where a primer set

CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary

CC to the complementary strand of a polynucleotide which comprises one of

CC the 5602 nucleotide sequences defined in the specification, where the

CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination

CC of an oligonucleotide comprising a sequence complementary to the

CC complementary strand of a polynucleotide which comprises a 5'-end

CC sequence and an oligonucleotide comprising a sequence complementary to a

CC polynucleotide which comprises a 3'-end sequence, where the

CC oligonucleotide comprises at least 15 nucleotides and the combination of

CC the 5'-end sequence/3'-end sequence is selected from those defined in

CC the specification. The primer sets can be used in antisense therapy and

CC in gene therapy. The primers are useful for synthesizing polynucleotides,

CC particularly full-length cDNAs. The primers are also useful for the

CC detection and/or diagnosis of the abnormality of the proteins encoded by

CC the full-length cDNAs. The primers allow obtaining of the full-length

CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and

CC AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to

CC AAH95893 represent human amino acid sequences; and AAH13629 to AAH13632

CC represent oligonucleotides, all of which are used in the exemplification

CC of the present invention.

XX Sequence 1736 BP; 449 A; 372 C; 453 G; 462 T; 0 other;

QY Query Match 43.8%; Score 602.8; DB 22; Length 1736;

Db Best Local Similarity 79.1%; Pred. No. 4.6e-178;

QY Matches 810; Conservative 0; Mismatches 167; Indels 47; Gaps 6;

Db	291	AGCAAGGAGCCTTAATGTTTAGTCCCAAGACCATGGGGAAGATGTCTCCAGACCAATGTCAG	350
Qy	631	AGGCTCTTCACAGCAGTCCCAATCAAACTACCTGGCCTGGAGGCCCTAGGAGAAAATGTTTTGT	690
Db	351	AGACCTTCACACAGCCCTCCTATACACAGGCCAGA----	405
Qy	591	GGGACAGCCCGAGGTCCTGTGCTGTGTGCAGCCTAGAGACTTGGTGCCCTGTGTGCCCA	750
Db	406	GGGCGTGTGCAGGGTCCGCTTGTCTGTGTGCAACTAGGAGTAGTGGCACCCTGTGTGCCCA	465
Qy	751	GTTAATTC-----AGCTGTGGCTTCAGAG	774
Db	466	GTCGCTTCACGCTGGCTGAAGGGGCCAATGTACAGCTCAGTTGTGGCCTCAGAG	525
Qy	775	GGTGAAGCCCCCAAGCCTTGGCAGCTTCCAAAGTGTGTGGAGCCTGTGGGTGCAAGAAG	834
Db	526	GGTGAAGCCCCAAGCTTTGGCAGGCTCCACATGCTGTGTGAGCCTGCGCAGTGCACAGAAG	585
Qy	835	TC--AAGAAATTGAGTTTGGGAACCTCCCAATCAGATTTTCAGAAAGATATATCGAAACCCCTG	893
Db	586	TCAAAGAATTGAGTTTGGGAACCTCTACTAGATTC-----AGATGTATGGAATGTCTA	642
Qy	894	GATGCCCGAGCAGAAGTTTGTCTGTAGGGGTGGGTTCCTCATCGAGAACCTCTCGCAAGGGT	953
Db	643	GATGCCCGAGCAAAAGCTTTGCTAT--GGGTGGGGCCCTCATAGAGAACCCTCTGCTAGGGC	701
Qy	954	AGTACAAAGGGGAATTTGGTGGGAGGCCGCCACACAGAGTCCCAAGTGGGGCTCCATC	1013
Db	702	AGTGTGGAAGGGAATGTGGGGTTGGAGCCCCACACAGAGTCCCTACTGGGGCACTGCC	761
Qy	1014	TAGTAGAGCTGTGCAGAAAGTCCACCATCTCTCCAGACTCCAGAAAGGTAGATCCACTGA	1073
Db	762	TAGTAGAGCTGTGCAGAAAGAGGCCATTGTCTTCTAGACCCCCAGAAATGGCAGACCCACTGA	821
Qy	1074	CAGCTTGCAGCATGTGCCCTGAAAAATCCACAGACACTCAGTGCAGCCTGTGAAAGCAGC	1133
Db	822	CAGCTTGCACCATGTCCCTGGAAAAAGTCGACACACTCAACGCCAGCCCGTGGAAAGCATC	881
Qy	1134	AGGATGGGAGTCTGACCTACAAAACCGTAGTGGCAGAGCTGACCAAGACCCGTGGGAAT	1193
Db	882	CGGAGGGAGGCTGTACCTGCGCAAGCCACAGGGGCAGAGTACCACAGACCATATGGGAAC	941
Qy	1194	CTACCTCTTGCATTGTTCATGACCTGGAGCTGAGACATGGAGTCAAAAAGAGATCATTTGG	1253
Db	942	TCACCTCTTGCATCAGTGTGACCTGSAITTTGAGACCTGSAATCAAGAGAGATCATTTGG	1001
Qy	1254	AGCTTTAAGATTTGACTGCCCACTGGATTTGCGACTTATATGAGGGGCCGT--ACCCCTTT	1312
Db	1002	AGCTTTAACATTTGACTGCCCACTGAATTTCTGGACTTGCATGGGAGCCCGTAACCCCAT	1061
Qy	1313	GTTTTGGCCAAATTTTTCATTTGGAATGCCGTAATTTACCCAATGCTGTACTCCATT	1372
Db	1062	GTTTTGGCCAAATTTCTCTCATTTGGAAATGGCTGTATTTTACCCAATACCTGCACACCCATT	1121
Qy	1373	GTAT 1376	
Db	1122	GTAT 1125	

Search completed: February 19, 2003, 23:48:32
Job time : 3592.57 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 19, 2003, 23:16:26 ; Search time 4390.26 Seconds
(without alignments)
9128.055 Million cell updates/sec

Title: US-09-997-610-1-copy_2_1378

Perfect score: 1377

Sequence: 1 atagtggtcatacctgtctt.....gctgtacctccattgtatg 1377

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl:

- 1: gb_ba:*
- 2: gb_htg:*
- 3: gb_in:*
- 4: gb_om:*
- 5: gb_ov:*
- 6: gb_pat:*
- 7: gb_ph:*
- 8: gb_pl:*
- 9: gb_pr:*
- 10: gb_ro:*
- 11: gb_sts:*
- 12: gb_sy:*
- 13: gb_un:*
- 14: gb_vi:*
- 15: em_ba:*
- 16: em_fun:*
- 17: em_hum:*
- 18: em_in:*
- 19: em_mu:*
- 20: em_om:*
- 21: em_or:*
- 22: em_ov:*
- 23: em_pat:*
- 24: em_ph:*
- 25: em_pi:*
- 26: em_ro:*
- 27: em_sts:*
- 28: em_un:*
- 29: em_vi:*
- 30: em_htg_hum:*
- 31: em_htg_inv:*
- 32: em_htg_other:*
- 33: em_htg_mus:*
- 34: em_htg_pln:*
- 35: em_htg_rod:*
- 36: em_htg_mam:*
- 37: em_htg_vrt:*
- 38: em_sy:*
- 39: em_htgo_hum:*
- 40: em_htgo_mus:*
- 41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	1281	93.0	145880	9	HS302D9	282198 Human DNA s
C 2	695.2	50.5	193092	9	AC017063	AC017063 Homo sapi
C 3	691	50.2	124518	9	AL138962	AL138962 Human DNA
4	690.8	50.2	77405	2	AL353634	AL353634 Homo sapi
5	690.2	50.1	40714	9	HSU212C1	Z69722 Human DNA s
6	689	50.0	123098	9	AC008799	AC008799 Homo sapi
7	689	50.0	154090	9	AC025577	AC025577 Homo sapi
8	689	50.0	174874	9	AC123595	AC123595 Homo sapi
9	689	50.0	186660	9	AC026107	AC026107 Homo sapi
C 10	689	50.0	189768	2	AC044889	AC044889 Homo sapi
C 11	684.2	49.7	170368	9	AC091982	AC091982 Homo sapi
12	683.6	49.6	67984	9	AC010312	AC010312 Homo sapi
C 13	683.2	49.6	168502	9	AC091005	AC091005 Homo sapi
14	681.6	49.5	131215	9	AC079614	AC079614 Homo sapi
C 15	680	49.4	73390	9	AC012038	AC012038 Homo sapi
C 16	679.8	49.4	152544	9	CNS05TEJ	AL359232 Human chr
C 17	679.8	49.4	203726	2	AC011882	AC011882 Homo sapi
18	679	49.3	173480	9	CNS00M8T	AL079343 Human chr
C 19	678.8	49.3	146743	2	AC093588	AC093588 Homo sapi
C 20	678.8	49.3	166679	9	AC079899	AC079899 Homo sapi
C 21	678.8	49.3	176426	9	AC007370	AC007370 Homo sapi
22	678.6	49.3	38235	9	AC004559	AC004559 Homo sapi
23	678	49.2	3417	9	AK091968	AK091968 Homo sapi
24	677.8	49.2	91323	9	AC106901	AC106901 Homo sapi
C 25	677.2	49.2	124531	9	HSJ633H17	AL049710 Human DNA
C 26	676.8	49.2	187898	2	AC016715	AC016715 Homo sapi
27	676.6	49.1	110000	2	AL732359_2	Continuation (3 of
C 28	676.6	49.1	138271	9	AC006360	AC006360 Homo sapi
29	676.6	49.1	147971	9	HS431P23	AL009178 Human DNA
C 30	676.4	49.1	77774	9	AP000339	AP000339 Homo sapi
31	676.4	49.1	100000	9	AP000217	AP000217 Homo sapi
C 32	676.4	49.1	340000	9	AP001760	AP001760 Homo sapi
33	675.8	49.1	107885	9	AC006389	AC006389 Homo sapi
34	675.8	49.1	150332	9	AC004921	AC004921 Homo sapi
C 35	675.6	49.1	181842	2	AL391823	AL391823 Homo sapi
36	675.4	49.0	94730	9	AP000230	AP000230 Homo sapi
37	675.4	49.0	100000	9	AP000144	AP000144 Homo sapi
38	675.4	49.0	100634	9	AP001594	AP001594 Homo sapi
39	675.4	49.0	340000	9	AP001695	AP001695 Homo sapi
C 40	675.2	49.0	155764	9	AC011238	AC011238 Homo sapi
C 41	675	49.0	182972	2	AC023550	AC023550 Homo sapi
C 42	674	48.9	174662	2	AC026036	AC026036 Homo sapi
43	673.8	48.9	123631	9	HS22F01	AL109967 Homo sapi
C 44	673.8	48.9	168863	9	AC011286	AC011286 Homo sapi
45	673.6	48.9	105692	9	AL451046	AL451046 Human DNA

ALIGNMENTS

RESULT 1
HS302D9
LOCUS
DEFINITION Human DNA sequence from clone RPI-302D9 on chromosome 22 Contains
GSs, complete sequence.
ACCESSION 282198
VERSION 282198.2 GI:6572207
KEYWORDS HTG.
SOURCE Homo sapiens.
ORGANISM Homo sapiens.
REFERENCE 1 (bases 1 to 145880)
AUTHORS Bridgeman,A.
TITLE Direct Submission

HS302D9 145880 bp DNA linear PRI 12-DEC-1999
Human DNA sequence from clone RPI-302D9 on chromosome 22 Contains
GSs, complete sequence.
282198
282198.2 GI:6572207
HTG.
Homo sapiens.
Homo sapiens.
Eukaryota: Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 145880)
Bridgeman,A.
Direct Submission

JOURNAL

Submitted (08-DEC-1999) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
 requests: clonerequest@sanger.ac.uk
 On Dec 13, 1999 this sequence version replaced gi:3164067.
 During difference assembly data is compared from overlapping clones.
 Where differences are found these are annotated as variations
 together with a note of the overlapping clone name. Note that the
 variation annotation may not be found in the sequence submission
 corresponding to the overlapping clone, as we submit sequences with
 only a small overlap as described above.
 This sequence has been finished according to sequence map criteria
 as follows. An attempt is made to resolve all sequencing problems,
 such as compressions and repeats, but not necessarily within known
 annotated human repeat sequence elements (e.g. Alu). Where the
 sequence is ambiguous, there is an annotation using the 'unsure'
 feature key.
 The following abbreviations are used to associate primary accession
 numbers given in the feature table with their source databases:
 Em.; EMBL; Sw.; SWISSPROT; Tr.; TREMBL; Wp.; WORMPEP; Information
 on the WORMPEP database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
 was generated from part of bacterial clone contigs of human
 chromosome 22, constructed by the Sanger Centre Chromosome 22
 Mapping Group. Further information can be found at
<http://www.sanger.ac.uk/HGP/Chr22>
 RPI-302D9 is from the library RPCI-1 constructed at the Roswell
 Park Cancer Institute by the group of Pieter de Jong. For further
 details see <http://bacpac.med.buffalo.edu/>
 VECTOR: PCYPAC2

This sequence is the entire insert of clone RPI-302D9 The true left
 end of clone CTA-282F2 is at 69682 in this sequence. The true right
 end of clone CTA-415G2 is at 55167 in this sequence.

FEATURES

source

1. 145880
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="22"
 /clone="RPI-302D9"
 /clone_lib="RPCI-1"

repeat_region
 188..245
 /note="MER3 repeat: matches 144..209 of consensus"
 repeat_region
 246..571
 /note="AluX repeat: matches 1..312 of consensus"
 repeat_region
 572..759
 /note="MER3 repeat: matches 1..144 of consensus"
 repeat_region
 783..933
 /note="MER5A repeat: matches 26..187 of consensus"
 repeat_region
 1033..1336
 /note="AluSp repeat: matches 1..299 of consensus"
 repeat_region
 1450..1583
 /note="MIR repeat: matches 24..160 of consensus"
 repeat_region
 1687..1752
 /note="L2 repeat: matches 2593..2661 of consensus"
 repeat_region
 2350..2660
 /note="AluSc repeat: matches 3..309 of consensus"
 repeat_region
 2684..2981
 /note="AluSq repeat: matches 2..300 of consensus"
 repeat_region
 3323..3343
 /note="MLTIE repeat: matches 116..136 of consensus"
 repeat_region
 3344..3652
 /note="AluY repeat: matches 1..309 of consensus"
 repeat_region
 3653..3928
 /note="MLTIE repeat: matches 136..359 of consensus"
 repeat_region
 3929..4278
 /note="THE1B repeat: matches 3..364 of consensus"
 repeat_region
 4279..4485
 /note="MLTIB repeat: matches 359..568 of consensus"
 repeat_region
 5073..5176
 /note="52 copies 2 mer ct 78 conserved"
 repeat_region
 5181..5491
 /note="AluJb repeat: matches 1..311 of consensus"
 repeat_region
 6369..6485
 /note="L2 repeat: matches 2579..2705 of consensus"

repeat_region
 6647..6685
 /note="MADE1 repeat: matches 1..23 of consensus"
 repeat_region
 6686..6987
 /note="AluX repeat: matches 1..302 of consensus"
 repeat_region
 6988..7036
 /note="MADE1 repeat: matches 23..77 of consensus"
 repeat_region
 7482..7754
 /note="AluJb repeat: matches 9..290 of consensus"
 repeat_region
 7775..8060
 /note="AluJo repeat: matches 1..295 of consensus"
 repeat_region
 8414..8551
 /note="L2 repeat: matches 2553..2706 of consensus"
 repeat_region
 8914..9030
 /note="MIR repeat: matches 147..262 of consensus"
 repeat_region
 9110..9280
 /note="MIR repeat: matches 91..262 of consensus"
 repeat_region
 9283..9412
 /note="MIR repeat: matches 15..144 of consensus"
 repeat_region
 9521..9679
 /note="FAM repeat: matches 3..161 of consensus"
 repeat_region
 9820..10225
 /note="MSTB repeat: matches 2..425 of consensus"
 misc_feature
 complement(10179..10678)
 /note="match: GSS: Em:B56592"
 misc_feature
 complement(10204..10728)
 /note="match: GSS: Em:AQ701486"
 misc_feature
 complement(10249..10706)
 /note="match: GSS: Em:AQ225495"
 repeat_region
 10312..10383
 /note="MIR repeat: matches 79..150 of consensus"
 misc_feature
 10718..11310
 /note="match: GSS: Em:B14024"
 misc_feature
 10784..11201
 /note="match: GSS: Em:B43656"
 repeat_region
 11838..11946
 /note="MIR repeat: matches 20..137 of consensus"
 repeat_region
 12174..12445
 /note="L2 repeat: matches 1988..2275 of consensus"
 repeat_region
 12444..12642
 /note="MIR repeat: matches 63..241 of consensus"
 misc_feature
 13017..13369
 /note="match: STS: Em:G49301"
 repeat_region
 13331..13397
 /note="MIR repeat: matches 174..244 of consensus"
 repeat_region
 13398..13698
 /note="AluSp repeat: matches 1..302 of consensus"
 repeat_region
 13699..13810
 /note="MIR repeat: matches 76..174 of consensus"
 repeat_region
 13806..13919
 /note="MIR repeat: matches 77..189 of consensus"
 repeat_region
 13945..14060
 /note="MIR repeat: matches 24..142 of consensus"
 repeat_region
 14061..14367
 /note="AluJ repeat: matches 1..301 of consensus"
 repeat_region
 14368..14452
 /note="MIR repeat: matches 141..225 of consensus"
 repeat_region
 14589..14679
 /note="MIR repeat: matches 173..262 of consensus"
 misc_feature
 14597..15201
 /note="match: GSS: Em:AQ553482"
 misc_feature
 14616..15060
 /note="match: GSS: Em:AQ370601"
 repeat_region
 14868..15040
 /note="MIR repeat: matches 49..233 of consensus"
 repeat_region
 15071..15188
 /note="L2 repeat: matches 2112..2239 of consensus"
 repeat_region
 15304..15399
 /note="MLTIB repeat: matches 1..99 of consensus"
 repeat_region
 15490..15662
 /note="AluSg1 repeat: matches 2..114 of consensus"
 repeat_region
 15669..15727
 /note="MLTIB repeat: matches 119..178 of consensus"
 repeat_region
 15728..16027


```

REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
1 (bases 1 to 193092)
Sulston,J.E. and Waterston,R.
Toward a complete human genome sequence
Genome Res. 8 (11), 1097-1108 (1998)
99063792
9847074
2 (bases 1 to 193092)
Boatright,E., Haakenson,W. and Doebber,A.
The sequence of Homo sapiens BAC clone RP11-354H17
Unpublished (2001)
3 (bases 1 to 193092)
Waterston,R.H.
Direct Submission
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
4 (bases 1 to 193092)
Waterston,R.H.
Direct Submission
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
5 (bases 1 to 193092)
Waterston,R.
Direct Submission
Submitted (27-MAR-2002) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
On Mar 15, 2002 this sequence version replaced gi:16596637.
-----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc
Contact: sapiens@watson.wustl.edu
-----
Summary Statistics
-----
Center project name: H_NH0354H17
-----

NOTICE: This sequence may not represent the entire insert of this
clone. It may be shorter because we only sequence overlapping
clone sections once, or longer because we provide a small overlap
between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded, sequenced with an alternate
chemistry, or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by sequence
from more than one subclone; and the assembly was confirmed by
restriction digest.

MAPPING INFORMATION:
Mapping information for this clone was provided by Dr. John D.
McPherson, Department of Genetics, Washington University, St. Louis
MO. For additional information about the map position of this
sequence, see http://genome.wustl.edu/gsc

SOURCE INFORMATION:
The RPCI-11 human BAC library was made from the blood of one male
donor, as described by Osoegawa,K., Woon,P.Y., Zhao,B., Frengen,E.,
Tateno,M., Catanese,J.J. and de Jong,P.J. (1998) An improved
approach for construction of bacterial artificial chromosome
libraries. Genomics 51:1-8. The clone may be obtained either from
Research Genetics, Inc. (http://www.resgen.com) or Pieter de Jong
and coworkers at http://www.chori.org
VECTOR: pBACE3.6

NEIGHBORING SEQUENCE INFORMATION:
The clone sequenced to the left is RP11-756P10, 2000 bp overlap;
the clone sequenced to the right is RP11-818C3. Actual start of
this clone is at base position 178381 of RP11-756P10; actual end is
at base position 193092 of RP11-354H17.

There is an unresolved tandem repeat from 23045 to 24207.

```

FEATURES
source

```

Location/Qualifiers
1..193092
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="UL"
/map="4"
/clone_lib="RPCI-11"
467..619
/rpt_family="L1"
repeat_region
840..880
/rpt_family="L1"
repeat_region
881..965
/rpt_family="(TG)n"
repeat_region
984..1039
/rpt_family="(TA)n"
repeat_region
1040..1071
/rpt_family="(TA)n"
repeat_region
1072..1210
/rpt_family="(TTTA)n"
repeat_region
1211..1478
/rpt_family="Alu"
repeat_region
1479..1490
/rpt_family="Alu"
repeat_region
2394..2715
/rpt_family="Alu"
repeat_region
3180..3470
/rpt_family="MaLR"
repeat_region
3891..3996
/rpt_family="Alu"
repeat_region
4022..4053
/rpt_family="MIR"
repeat_region
4291..4417
/rpt_family="AT_rich"
repeat_region
4830..4999
/rpt_family="MIR"
repeat_region
6479..6507
/rpt_family="MER1_type"
repeat_region
6691..6989
/rpt_family="AT_rich"
repeat_region
7734..7843
/rpt_family="Alu"
repeat_region
9245..9274
/rpt_family="L1"
repeat_region
10159..10408
/rpt_family="AT_rich"
repeat_region
10793..10968
/rpt_family="MIR"
repeat_region
10982..11144
/rpt_family="(TATAA)n"
repeat_region
11157..11193
/rpt_family="(TA)n"
repeat_region
12999..13157
/rpt_family="(TG)n"
repeat_region
13192..13238
/rpt_family="L2"
repeat_region
13297..14044
/rpt_family="AT_rich"
repeat_region
14386..14429
/rpt_family="L1"
repeat_region
14565..15340
/rpt_family="MIR"
repeat_region
15608..16083
/rpt_family="L1"
repeat_region
16094..16267
/rpt_family="MaLR"
repeat_region
17080..17178
/rpt_family="MIR"
repeat_region
17211..17235
/rpt_family="GA-rich"
repeat_region
17237..17522
/rpt_family="AT_rich"
repeat_region
17581..17975
/rpt_family="Alu"
repeat_region

```


Dye-terminator Big Dye; 86% of reads
Consensus quality: 7604 bases at least Q40
Consensus quality: 7708 bases at least Q30
Consensus quality: 7730 bases at least Q20
Insert size: 77405; sum-of-contigs
Insert size: 85298; 1.5% error; agarose-fp
Quality coverage: 4.54x in Q20 bases; sum-of-contigs Quality
coverage: 4.30x in Q20 bases; agarose-fp

* NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.

FEATURES
Location/Qualifiers
source 1..77405
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="X"
/map="q21.33-22.3"
/clone="RP11-435A2"
/clone_lib="RPC1-11.2"
misc_feature 1..77405
/note="assembly_fragment:00554"

BASE COUNT 25072 a 14551 c 15133 g 22649 t
ORIGIN

Query Match 50.2%; Score 690.8; DB 2; Length 77405;
Best Local Similarity 80.6%; Pred No. 1.1e-193;
Matches 871; Conservative 0; Mismatches 172; Indels 37; Gaps

QY 334 CTGCATCATTTGCAAGTGTAATATTTGGCTTAATGAGGAAGCAAAATTTGGCTAATAAGGAA 393
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 8637 CTGCCCTAGAGATCTCGGAATTTGACCTTGAGAGAGATGATTTAGGCTATCAGGTGGA 8696
QY 394 GAAATTTCTAAGCAGCAAGACATTCAGAGGTGACTTGGTGGCTGCTTTAAAGCATTCAGT 453
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 8697 GAAATTTCTAAGCAGCAAGACATTCAGAGGTGACTTGGTGGCTGCTTTAAAGGCATTCAGT 8756
QY 454 TTCAATAGGGAGGCAGACGATAGAGCTTCAGAAAAATTTGCACCCCTGACAAATGTGATAAA 513
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 8757 TTTGCAAGGACAGCAGACATAAAATTCAGAAATTTTCAGCCCTGCACANTGTGATGNA 8816
QY 514 AAGAAAAACCCA-TTTTCTGAGGGGAAATTCAGCTGGCTGCAGAAATTTGCATATGTAA 572
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 8817 AAGAAAAACCCATTTTCTGATGATAAATTAAGCCAGCTGCAGAAATTTGCATAAGTAA 8876
QY 573 TGAGGAGCTGAATGTTAATCCTCAAGACAAATGGGGAAAAATATCTCTCGACATGTCAAG 632
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 8877 CGAGGAGCCAAATGTTAATCACTAAGACATCGGGAAAAATATCTCATGGCTTGTCTCAG 8936
QY 633 GTCTTCACAGCAGCTCAATCAATCACTGG--CCTGGAGGCTTAGGAGAAATGGTTTTGT 690
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 8937 GTCTTCAGGCAGCCCTCACATCACTAGCCGCCCCAGAGGCCTTAGAGAAATGGTTTTAT 8996
QY 691 GGGACAGGCCAGGGTCCCTGTGTGTGTGCAGCCTTAGAGACTTGGTGCCTGTGTCCCA 750
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 8997 GGGTGGGCCAGGGTCCCATGCTATGTGCAGTCTAGGAGCTTGGTGCCTGTCCTGTC 9056
QY 751 GTTAATTCA-----GCTGTGGCTTCAGAGGT 777
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 9057 GCCACTCCAGCCATAACTGAAAGGGGCCAAGGTATAGCTTGGGCTGTTCCTCAAGGAT 9116
QY 778 GCAAGCCCCAAGCTTTGGCAGCTTCCCAAGTGTGTGTAGCCCTGTGGGTGCCAAGAGTCA 837
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 9117 GGAAGCCCCAGGCCTTGGCAGCTTCCATGTGTGTGTGTAGCCCTGCAGGTGCATGAAGTCA 9176
QY 838 AGAATTCAGGTTTTGGGAACCTCCAAATTCAGATTTTCAGAGATATATGGAACCCCTGGATG 897
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 9177 AGAATTCAGGTTTTGGGAACCTCCACCTAGATTCAGAGATGATGGAATGCCCTGGATG 9236
QY 898 CCCAGGCAGAAAGTTTGCTGTAGGGGTGGGGTCCCTCATGGAGAACCTCTGCAAGGGTAGTA 957
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 9237 CCCAGGCAAAAGTTTGCTGCAGGGGCGAGCCTCATGGAGAACCTCTGTTAGGGTGGT 9296
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Qy	958	CAAAAGGAAATGTTGGGTGGAGCGCCCCACACAGAGTCCCAAGTGGGGCTCCATCTAGT	1017
Db	9297	TGGAAGGGGAAAGTGGGGTCTATAGCGGCCACACAGAGTCCCTACCGGGGACCACTAGT	9356
Qy	1018	AGAGCTGTGAGAAAGTCCACCACCTCTCCAGACTCCCAAGAGGTAGATCCACATGCACAGC	1077
Db	9357	GGAGCTGTGAGAAAGGGGCCACCACTCTCCAGACCCCAAGATGTTAGATCCACCGACAGC	9416
Qy	1078	TTGCAGCATGTGCCTGAAAAATCCACACACACTCAGTCCCAAGCCTGTGAAAGCAGCAGGG	1137
Db	9417	CTGCACCCTGTGCCTGGAAAAGCCGACACACTCAATGCCAGCCTGTGAAAGCAGCAGGG	9476
Qy	1138	ATGGAGTCTGTACCCCTACAAAACCGTAGTGGCAGAGTGCACACAGACGCTGGGAATCTAC	1197
Db	9477	AGGGAGGCTGTACCTGTCAAGAGCCACAGGGGCAGAGTGCCTCAAGATCATGGGAACCAAT	9536
Qy	1198	CTCTTGCAATCTCATGACCTCGAGCTGAGACATCGAGTCAAAAGACATCATTTGGAGCT	1257
Db	9537	CTCTTGCAATGACATGACCTCGATGATGAGACATGGAGTCAAAAGAAATCATTTGGAGCT	9596
Qy	1258	TTAAGATTGACTGCCCCACTTGGATTTGGGACTTATATGGGGCCCGTA-CGCCCTTTGTTT	1316
Db	9597	TTAAGATTGACTGCCCACTGGATCTCAGACTTCCATGGGCGCTGTAGCCTCTTTGTTT	9656
Qy	1317	TGGCAATTTTTCGATTTGGAATGCCGATTCGCTATTTACCCATGCTGTACCTCCATGTAT	1376
Db	9657	TGGCAATATCTCCATTTGGAATGGCTGTATTTATCCAATGCCGTGACCCCATTTGTAT	9716
RESULT 5			
LOCUS	HSU212C1	40714 bp	DNA linear PRI 23-NOV-1999
DEFINITION	Human DNA sequence from cosmid u212c1, between markers DXS366 and DXS87 on chromosome X.		
ACCESSION	Z69722		
VERSION	Z69722.1	GI:1204119	
KEYWORDS	X.		
SOURCE	Human sapiens.		
ORGANISM	Human sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
AUTHORS	Odell,C.		
TITLE	Direct Submission		
JOURNAL	CB10 1RQ, UK. E-mail enquires: humpbesanger.ac.uk		
COMMENT	IMPORTANT: This sequence is the entire insert of clone U212C1. The true left end of clone U212C1 is at 1 in this sequence. The true right end of clone U212C1 is at 40714.		
FEATURES	U212C1 is from the human chromosome X-specific cosmid library.		
source	Location/Qualifiers		
	1. .40714		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
	/chromosome="X"		
	/map="X"		
repeat_region	/clone="LL0XNC01-212C1"		
	/clone_lib="LL0XNC01"		
	10. .233		
repeat_region	/note="L1 element fragment"		
	388. .758		
repeat_region	/note="L1 element fragment"		
	932. .1108		
repeat_region	/note="L1 element fragment"		
	2571. .2617		
repeat_region	/note="MER43 element fragment"		
	2613. .5203		
repeat_region	/note="L1 element fragment"		
	5204. .5229		
repeat_region	/note="13 copies of 2 mer 100 % conserved"		
	5230. .5427		
repeat_region	/note="MER43 element fragment"		

AC025577
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM

AC025577.15 GI:9910028
 HTG.
 Homo sapiens.
 Homo sapiens
 Mammalia; Chordata; Craniata; Vertebrata; Euteleostomi;
 1 (bases 1 to 154090)

REFERENCE
 AUTHORS

Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C., Alsbrooks,S.L., Amarantunge,H.C., Are,J.R., Banks,T., Barbaria,J., Benton,J., Binage,K., Blankenburg,K., Bonnin,D., Bouck,J., Burche,S., Brileva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C., Burch,P., Burrell,C., Burrell,K.L., Byrd,N.C., Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Z., Chiu,D., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Emerling,S., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K., Han,J., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J., Hernandez,O., Hodgson,A., Hogue,M., Holloway,C., Hollins,B., Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Ioshikhes,I., Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lee,E., Lewis,L.C., Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Louisghe,H., Lozano,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J., Maheshwari,M., Mapua,P., Marondel,I., Martin,R., Martindale,A., Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Merscher,S., Metzker,M., Miller,A., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Montgomery,K.T., Morgan,M., Morris,S., Moser,M., Muzny,D., Neal,D., Nelson,D., Newton,J., Newton,S., Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokendo,S., Ogih,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L., Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojebokan,I., Rolfe,M., Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shm,C., Shoshitari,N., Sisson,I., Sodergren,E., Sonalet,T., Sparks,A., Stanley,H., Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamrlisa,A., Tamerisa,K., Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S., Usmani,K., Vasquez,L., Veta,V., Villalon,D., Vinson,R., Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S., Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Kucherlapati,R., Nelson,D. and Gibbs,R.

Direct Submission
 Unpublished
 2 (bases 1 to 154090)
 Worley,K.C.
 Direct Submission
 Submitted (11-MAR-2000) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
 3 (bases 1 to 154090)
 Worley,K.C.
 Direct Submission
 Submitted (25-AUG-2000) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
 4 (bases 1 to 154090)
 Worley,K.C.
 Direct Submission
 Submitted (30-APR-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
 On Aug 25, 2000 this sequence version replaced gi:9664948.
 INFORMATION: <http://www.hgsc.bcm.tmc.edu/> or email gc-help@bcm.tmc.edu

CLONE LENGTH: This sequence does not necessarily represent the entire insert of this clone. Overlapping regions of clones are only sequenced and submitted once, so the sequence for the remainder of the insert may be found in the record for the adjacent clones. Overlapping clones are noted at the beginning and end of the Features listing.

ANNOTATION OF FEATURES:

STSS are identified using ePCR (Genome Res. 7:541-550) searches of a local database that includes entries from dbSTS, GDB, and local mapping efforts.

Repeats are identified using RepeatMasker (A. Smit and P. Green, unpublished.) for Human and Mouse sequences.

Genes and Region of sequence similarity are identified by BLAST (Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the EST and cDNA sequences. Genes demonstrate at least two exons flanked by consensus splice sites that maintained sequence continuity across the splice junctions. Sequences that are not identical matches are annotated as similar.

SEQUENCING READ COVERAGE: Sequencing is completed to a minimum standard of double strand coverage with a minimum of 2 clones and 2 reads with no ambiguities or 2 chemistries with a minimum of 2 clones and 3 reads with no ambiguities. If the sequence quality for a region does not meet this standard, it will be indicated in the annotation as Low Coverage.

QUALITY OF INDIVIDUAL BASES: This sequence meets stringent quality standards - estimated error rate less than 1 per 10,000 bases. Reports of low quality individual bases and measures of base quality are listed below. Description of the metrics can be found at URL: <http://gc.bcm.tmc.edu:8088/quality.info/genbank.annotation.html>.

QUALSTAT-REPORT-----

----- Summary Statistics -----
 Contig length: 154090
 Phrap values in estimate: 153345
 Average error rate (BCM-Phrap estimate): 0.000261755
 Fraction of Phrap values less than 40 : 0.0284086
 Number of consensus changing edits: 22
 Number of N's in consensus : 0

----- Consensus changing edits -----

Position	Original+Context	Edited+Context
9347	aacacagacc(n)tttttttttt	aacacagacc(t)tttttttttt
10135	atttgcacct(n)taggaaaaga	atttgcacct(t)taggaaaaga
10353	ctgtcacacg(n)aaagatcca	ctgtcacacg(t)aaagatcca
10417	agcaatcgag(n)ggctacagaa	agcaatcgag(t)ggctacagaa
61959	cittactatt(n)tgtttatctt	cittactatt(t)tgtttatctt
61960	ttactatt(n)gtttattctt	ttactatt(t)gtttattctt
61961	ttactatt(n)gtttattctt	ttactatt(t)gtttattctt
61982	acaacacag(n)taggttttgg	acaacacag(a)taggttttgg
61983	caacacag(n)taggttttgg	caacacag(a)taggttttgg
61993	ttaggtttgg(n)ttactcttcc	ttaggtttgg(t)ttactcttcc
62205	tcattgacct(n)ctggctattc	tcattgacct(c)ctggctattc
62956	actgcaacct(n)tgctccccag	actgcaacct(c)tgctccccag
63722	ttcacatata(n)cagctacagta	ttcacatata(a)cagctacagta
90212	cttagaaaaa(n)gacnttttct	cttagaaaaa(t)gacnttttct
90216	gaaaaangac(n)tttctttttt	gaaaaangac(t)tttctttttt
90228	tttcttttta(n)ttttttonat	tttcttttta(t)ttttttonat
90236	taattttttc(n)atctcttaggg	taattttttc(t)atctcttaggg
90443	accttagatg(n)ctctccccag	accttagatg(a)ctctccccag
91405	cacagcttaa(n)caaggttaaga	cacagcttaa(a)caaggttaaga
122290	aaaaaagaaa(n)gtaagagaaa	aaaaaagaaa(a)gtaagagaaa
137090	acagagaaag(n)caaaacactc	acagagaaag(a)caaaacactc
141632	tgaagccact(n)gaaaagtaat	tgaagccact(t)gaaaagtaat

----- Distribution of Quality < 40 Bases -----


```
1000| * * * * *
900| * * * * *
800| * * * * *
700| * * * * *
600| * * * * *
500| * * * * *
400| * * * * *
300| * * * * *
200| * * * * *
100| * * * * *
0| * * * * *
```

Phrap Value Range

5 10 15 20 25 30 35 40

```
Version: 1.01 gxf.
FEATURES
  source
    Location/Qualifiers
      1..154090
        /organism="Homo sapiens"
        /db_xref="taxon:9606"
        /chromosome="12"
        /clone="RP11-13C3"
      1..1231
        /note="overlaps bases 14771..16001 of clone AC079952"
        /function="clone overlap"
      54..288
        /rpt_family="MIR"
      1020..1146
        /rpt_family="MER3"
      complement(1248..1584)
        /rpt_family="MER46B"
      complement(1624..1682)
        /rpt_family="MIR"
      2630..2761
        /rpt_family="GA-rich"
      4088..4216
        /rpt_family="L2"
      5232..5301
        /rpt_family="MER102"
      complement(5450..5749)
        /rpt_family="AluY"
      6571..6616
        /rpt_family="(TG)n"
      complement(6678..7320)
        /rpt_family="LIME3"
      7419..7454
        /rpt_family="AT-rich"
      complement(7741..7807)

Query Match      50.0%; Score 689; DB 9; Length 154090;
Best Local Similarity 80.1%; Pred. No. 4e-193;
Matches 850; Conservative 0; Mismatches 175; Indels 36; Gaps 2;

QY 351 GAATATTGGGCTAATGAGGACCAATTTTGGCTAATAAGGAAGAAATTTCTAAGCAGCA 410
Db 101594 GAACTTTGACATTGAGACATGATTTAGGGTATCTGGGGAAGAAATTTCTAAGTAGCA 101653

QY 411 AAGCATTTCAAGAGGTGACTTGGGTGCTGTTAAAGGCATTACGTTTCATAAAGGGAGGCAGA 470
Db 101654 AAGCATTTAAAGAGGTGACTTAGATGTTTAAAGGTATTTCAGTTATATAAGGGAGGCAGA 101713

QY 471 GCATAGAGTTTCAGAAATTTCCACCTGCACAAATGTGATATAAAAGAAACCCATTTC 530
Db 101714 GCATAAACCTTTGGAAATTTTCAGCCTGCACAAATGTGATATAAAAGAAACCCATTTC 101773

QY 531 TCAGGGGAATTTCAAGCTGGCTGCAGAAATTTGTCATATGTAATGAGGAGCTCAATGTTAA 590
Db 101774 TCAGGGGAATTTGAAGCTGGCTGCAGAAATTTGTCATATGTAATGAGGAGCTCAATGTTAA 101833

QY 591 TCCTCAAGACAATGGGGAATATCTCTGGACATGTCTGAGAGGTCTTCACAGCAGTCCAT 650
```

```
Db 101834 TCCCCAAGACAATGGGGAATATGTCTCCAAGGCATATCAGAGTCTTCTCATGCGAGCCAT 101893
QY 651 CAATACACTGGCTGGAGGCTTAGGAGAAATGGTTTCTGGGACAGGCCAGGCTTCCCT 710
Db 101894 CCCATCAGAGGCCGAGGCTTAGGAGAAATGGTTTCTAGGCCAGGCCAGGAGCCCC 101953
QY 711 GTGCTGTGTGCAGCCTTAGAGACTTGGTGCCCTGTGT- 746
Db 101954 ATGCTGTGTGCAGCCTTAGGACTTGGTGCCCTGTGTATACAGCTGCTGCTGAGTGGGCT 102013
QY 747 -----CCAGTTAAATTCAGCTGTGGCTTCAGAGGGTGCAAGGCCCAAGCCCTTGG 795
Db 102014 GAAAGGGCCAGCATAGAGCTTGGGCGGTGGCTTCGAGGGTGCAAGGCCCAAGCCCTTGG 102073
QY 796 CAGCTTCCCAAGTGTGTGAGCCTGTGGGTGCAGAAAGTCAAGAAATTCAGGTTTGGGAA 855
Db 102074 CAATTCCCATGTGTGTGAGCCTGAGGTGCAGAAAGTCAAGAAATTCAGGTTTGGGAG 102133
QY 856 COTCCAATCAGATTTTCAGAAAGATATATGAAACCCCTGGATGCCAGGCAGAGTTTGTCT 915
Db 102134 COTCCACCTTAGATTTTCAGAGGATGTATGGAATGCTGATGTCCAGGCAGAGTTTGTCT 102193
QY 916 GTAGGGGTGGGTCTCTCATGTGAGAAACCTCTGCAAGGGTAGTACAAAGGAAATGTGGG 975
Db 102194 TCAGGGGTGGGCGCTCATGTGAGAAACCTCTGCTAGGGCAGTGCAGAAAGGAAATGTGGG 102253
QY 976 TGGGAGCCCCACAGAGTCCCGAGTGGGCTCCATCTAGTAGAGCTGTGAGAGAACT 1035
Db 102254 TTGGAGCCCATACAGAGTCTCTTCTGTTGGGACCATCTAGGGGAGCTGTGAGAGAGG 102313
QY 1036 CCACCATCTCCAGACTCCAGAAAGGTAGATCCACTGACAGCTTGCAGCATGTGCCTGAA 1095
Db 102314 CCACCATCTCCAGACCCAGAAATGATAGATCCACCAACAGCTTGACCATGTACCTGGA 102373
QY 1096 AAATCCACAGACTCAGTGCCAGCTGTGAAAGCAGCAGGATGGAGTCTGTACCTAC 1155
Db 102374 AAAGCCACAGAGACTCAACACCGCCCTGAAAGCAGCCAGGAGGAGCTGTACCTGC 102433
QY 1156 AAACCGTAGTGCAGAGCTGACCAAGCCGTGGGAATCTACCTCTTGCATGTCATGAC 1215
Db 102434 AAAGCCACAGAGCAGAGCTTCCCAAGACCTGGGAACCCACTCTGTGAGAGTGAC 102493
QY 1216 CTGGAGCTGAGACATGGAGTCAAAAGAGATCATTTTGGAGCTTTAAGATTTGACTGCC 1275
Db 102494 CTAGATGTGAACATGCAGTCAAGAGATCATATTAGAACTTTGAGATTTGACTACCT 102553
QY 1276 ACTGATTTCCGACTTATATGAGGCGCCGTA-CCCCCTTTGTTTGGCCAAATTTTCCATT 1334
Db 102554 TCTGGGTTTCAGACTTGCATGGGCAATAGCCCTTTGTTTGGCCAAATTTTCCCAAT 102613
QY 1335 TGGAACTCCGTTATTTACCAATGCTGTACCTCCATTGTA 1375
Db 102614 TAGAACGCTGTATTTACTCAATGCTGTATCGCCATTGCA 102654

RESULT 8
AC123595
LOCUS AC123595 Homo sapiens chromosome 5 clone RP11-213H15, complete sequence.
DEFINITION AC123595
ACCESSION AC123595.1 GI:21306655
VERSION AC123595.1
KEYWORDS HTG
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 174874)
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE Direct Submission
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 174874)
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.
```

TITLE		Direct Submission	
JOURNAL	Submitted (01-JUN-2002) DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA		
COMMENT	Draft Sequence Produced by DOE Joint Genome Institute www.jgi.doe.gov		
Finishing Completed at Stanford Human Genome Center www.shgc.stanford.edu			
Quality: Phrap Quality >=40 100% of Sequence;			
Estimated Total Number of Errors is 0.			
Location/Qualifiers			
FEATURES	1. .174874		
source	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
	/chromosome="5"		
	/clone="RP11-213H15"		
BASE COUNT	52170 a 35353 c 35890 g 51461 t		
ORIGIN			
Query Match 50.0%; Score 689; DB 9; Length 174874;			
Best Local Similarity 80.1%; Pred. No. 4e-193;			
Matches 850; Conservative 0; Mismatches 175; Indels 36; Gaps 2;			
Qy	351	GAATATTGGCTAATGAGGACCAATTTTGGCTAATAAGAGAAATTTCTAAGCAGCA	410
Db	54218	GAATTTTAACTTGAAGAAGATGATTTAGGTGATCTGGCAGAGAATTTCTAAGCAGCA	54277
Qy	411	AAGCATTCAAGAGTGACTTGGGTGCTGTTAAAGGCATTCAGTTTCAAGGGAGGCAGA	470
Db	54278	AAGCATTCCAGAGTGACTTGGGTGCTGTAAAGGTATTGATTTTAAGGGAGGCAGA	54337
Qy	471	GCATAAGATTCAGAAAATTTGCAACCTGCAATGTGATATAAAGAAAAACCCATTTTC	530
Db	54338	GCATAAAGCTCGAAAAATTTGCAAGTCGCAATGTGATAGAAAAGAAAAATTCATTTTC	54397
Qy	531	TGAGGGAAATTCAGCTGGCTGCAGAAATTTGCATATGTAATCAGGAGCTGATGTAA	590
Db	54398	TGAGGAGAAATTCAGCTGGCTGCAGAAATTTGCATATGTAATCAGGAGCTGATGTAA	54457
Qy	591	TCCTCAAGACAAATGGGAAATATCTCTCGTGCATGTCCAGAGTCTTCACAGCAGTCCAT	650
Db	54458	TCCCCAAGACAAATGGGAAACATCTCCAGGCATGTCCAGAGTCTTCACGGCAGCCCT	54517
Qy	651	CAATCTACTGGCTGGAGGCTTAGGAGAAA ---TGTTTTTTGGGACAGGCCAGGGTC	707
Db	54518	CCCATCATAAACCCAAAGGCTAGGAGGAAAGTGGTTTTGTGGCCAGGCCAGGGTC	54577
Qy	708	CCTGTGCTGTGCAGCCTAGAGACTTGGTCCCTGTGCCAGTTAATCA -----	759
Db	54578	CTTGTGCTGTGCAGCTTAGGACTTGGTCCCTGTGCCAGTCCAGCCTGCTCAGGCATGGC	54637
Qy	760	-----GCTGTGGCTTCAGAGGGTCAAGGCCCAAGCCCTTG	794
Db	54638	TGAAGGGGCCAACATAGAGCTCGGGCTGTGGCTTCAGAGGTGCAGAGCCCAAGCCTTG	54697
Qy	795	GCAGCTTCCAAGTGGTGTGAGCCTGTGGGTGCAAGAGTCAAGAAATTCAGGTTGGGA	854
Db	54698	GCAGCTTCCATGTGGTGTGAGCCTGCAAGTGCACAGAGTCACAAAATTTGGGTTTGGGA	54757
Qy	855	ACCTCCATCAGATTCAGAGATATATGGAACCCCTGGATGCCAGGCAGAGAGTTTGC	914
Db	54758	ACCTCCACCTAGATTGAGATGATGTATGGAATATGCCCGATGACAGGAGAGTTTGC	54817
Qy	915	TGTAGGGTGGGTCCCTCATGGAGAACTCTGCAAGGGTAGTACAAAAGGAAATGTTGG	974
Db	54818	TGCAGGGCGAGGCTTCAATGGAGAACTCTGTAGGGCAGTCGAGAGAAAAATGTAGG	54877
Qy	975	GTGGAGGCCCCACACAGAGTCCCGAGTGGGGCTCCCATCTAGTAGAGCTGTGAGAAGAAG	1034
Db	54878	GTCAGAGCCACACAGAGATACCCTAGACACTGCCTAGTGAGCTGTGAGAAGAAG	54937
Qy	1035	TCCACCATCTCCAGACTCCAGAGGGGTAGATCCATGACAGCTTGCAGCATGTGCCTGA	1094
Db	54938	GCCACATCTCCAGACTCCAGAAATGTTAGTGTCCACCAACAGCTTGACCTGTGACCTGG	54997
Qy	1095	AAAATCCACAGACACTGTCGAGCCTGTGAAAGCAGCAGGATGGAGTCTGTACCCCTA	1154
Db	54998	AAAAGCCACAGGCACCTCATTTGCCAGCCTGTGAAAGCAGCAGGAGGAGCTATCCCTG	55057
Qy	1155	CARAAACCGTAGTGGCAGAGCTGACCAAGACCGTGGGAATCTACCTCTTGATTCATGA	1214
Db	55058	CAAAGTCACAGGGCGAAGCTGTCAGAGCCTTGGGAACCTACTCTTGTGCTCAGCGTGA	55117
Qy	1215	CCTGGAGCTCAGACATGGAGTCAAAAGAGATCATTTTGGAGCTTTAAGATTTGACTGCC	1274
Db	55118	CCCGATGTCAGACATGGAGTCAAAAGAGATCATTTTGGAGCTTTAAGATTTGACTGCC	55177
Qy	1275	CACTGGATTTCCGACTTATATGGGGCCGCTACCCCTTTGTTTTGGCCAAATTTTTCAT	1334
Db	55178	TGCTGGATTTCCAGACTGCATGGGGCCTGTAGCCCTTTGCTTTGGCCAAATTTTTCAT	55237
Qy	1335	TGGAAGCTCGCTATTTACCCCAATGCCTGTACTCCATTGTA	1375
Db	55238	TGGAACAGCTGTATTTACCCCATGCCTGTACCCCGATTGTA	55278
RESULT 9			
AC026107	Homo sapiens chromosome 12 clone RP11-307L1, WORKING DRAFT		
LOCUS	SEQUENCE, 1 unordered piece.		
DEFINITION	AC026107.22 GI:13899178		
ACCESSION	HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.		
VERSION	HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.		
KEYWORDS	Homo sapiens		
SOURCE	Homo sapiens		
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
REFERENCE	1 (bases 1 to 186660)		
AUTHORS	Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-ouman,F.R., Allen,C., Alsbrooks,S.L., Amaratunga,H.C., Are,J.R., Ayale,M., Banks,T., Barbara,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D., Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Caron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C., Coyte,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Haves,A., Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B., Homsfi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loulseged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenko,S., Ogum,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Renvy,G., Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G., Scher,S., Scott,G., Shen,H., Shoohtari,N., Sisson,I., Sodergren,E., Sotaike,T., Sparks,A., Stanley,H., Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H., Tansay,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalob,D., Vinson,R., Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S., Williams,G., Williamson,A., Wleczky,R., Woodson,S., Worley,K.,		

AUTHORS

Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N., Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F., Boguslavsky,L., Boukhgalter,B., Brown,A., Burkett,G., Campopiano,A., Castile,A., Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L., Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L., Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J., Lacroque,K., Lamazares,R., Landers,T., Lehoczkzy,J., Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N., McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R., Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J., Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnelli,P., O'Neill,D., Oliver,T.M., Oliver,J., Peterson,K., Pierre,N., Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zimmer,A. and Zody,M.

TITLE
JOURNAL

Submitted (12-APR-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

3 (bases 1 to 189768)

REFERENCE
AUTHORS

Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N., Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F., Boguslavsky,L., Boukhgalter,B., Brown,A., Burkett,G., Campopiano,A., Castile,A., Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L., Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L., Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J., Lacroque,K., Lamazares,R., Landers,T., Lehoczkzy,J., Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N., McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R., Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J., Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnelli,P., O'Neill,D., Oliver,T.M., Oliver,J., Peterson,K., Pierre,N., Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zimmer,A. and Zody,M.

TITLE
JOURNAL

Submitted (24-AUG-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

On May 22, 2000 this sequence version replaced gi:7543856.

COMMENT

All repeats were identified using RepeatMasker:

Smit, A.F.A. & Green, P. (1996-1997)

<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: <http://www-seq.wi.mit.edu>

Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information

Center project name: L9594

Center clone name: 792_C_1

----- Summary Statistics

Sequencing vector: M13; M7815; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.960731

Consensus quality: 168749 bases at least Q40

Consensus quality: 180047 bases at least Q30

Consensus quality: 184014 bases at least Q20

Insert size: 190000; acarose-fp

Insert size: 186268; sum-of-contigs

Quality coverage: 3.8 in Q20 bases; agarose-fp

Quality coverage: 3.9 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
 * consists of 36 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

* 1 1015: contig of 1015 bp in length
 * 1016 1115: gap of 100 bp
 * 1116 2284: contig of 1169 bp in length
 * 2285 2384: gap of 100 bp
 * 2385 3898: contig of 1514 bp in length
 * 3899 3998: gap of 100 bp
 * 3999 5259: contig of 1261 bp in length
 * 5260 5359: gap of 100 bp
 * 5360 6912: contig of 1553 bp in length
 * 6913 7012: gap of 100 bp
 * 7013 8145: contig of 1133 bp in length
 * 8146 8245: gap of 100 bp
 * 8246 9694: contig of 1449 bp in length
 * 9695 9794: gap of 100 bp
 * 9795 11004: contig of 1210 bp in length
 * 11005 11104: gap of 100 bp
 * 11105 12907: contig of 1803 bp in length
 * 12908 13007: gap of 100 bp
 * 13008 14916: contig of 1909 bp in length
 * 14917 15016: gap of 100 bp
 * 15017 17114: contig of 2098 bp in length
 * 17115 17214: gap of 100 bp
 * 17215 18527: contig of 1313 bp in length
 * 18528 18627: gap of 100 bp
 * 18628 20561: contig of 1934 bp in length
 * 20562 20661: gap of 100 bp
 * 20662 22484: contig of 1823 bp in length
 * 22485 22584: gap of 100 bp
 * 22585 25287: contig of 2703 bp in length
 * 25288 25387: gap of 100 bp
 * 25388 27603: contig of 2216 bp in length
 * 27604 27703: gap of 100 bp
 * 27704 30020: contig of 2317 bp in length
 * 30021 30120: gap of 100 bp
 * 30121 33472: contig of 3352 bp in length
 * 33473 33572: gap of 100 bp
 * 33573 38084: contig of 4512 bp in length
 * 38085 38184: gap of 100 bp
 * 38185 41069: contig of 2885 bp in length
 * 41070 41169: gap of 100 bp
 * 41170 45266: contig of 4097 bp in length
 * 45267 45366: gap of 100 bp
 * 45367 51092: contig of 5726 bp in length
 * 51093 51192: gap of 100 bp
 * 51193 55262: contig of 4070 bp in length
 * 55263 55362: gap of 100 bp
 * 55363 60982: contig of 5620 bp in length
 * 60983 61082: gap of 100 bp
 * 61083 67628: contig of 6546 bp in length
 * 67629 67728: gap of 100 bp
 * 67729 72690: contig of 4962 bp in length
 * 72691 72790: gap of 100 bp
 * 72791 79286: contig of 6496 bp in length
 * 79287 79386: gap of 100 bp
 * 79387 83451: contig of 4065 bp in length
 * 83452 83551: gap of 100 bp
 * 83552 92170: contig of 8619 bp in length
 * 92171 92270: gap of 100 bp
 * 92271 96333: contig of 4063 bp in length
 * 96334 96433: gap of 100 bp
 * 96434 103218: contig of 6785 bp in length
 * 103219 103318: gap of 100 bp
 * 103319 112553: contig of 9235 bp in length
 * 112554 112653: gap of 100 bp
 * 112654 123239: contig of 10586 bp in length

www-shgc.stanford.edu
Quality: Phrap Quality >=40 99.8% of Sequence;
Estimated Total Number of Errors is 0.2.
STS Content:
SHGC-10855 G13672
SHGC-53350 G36743
WI-14854 G22310
SHGC-64132 G38710.

FEATURES
source
1. .170368
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="5"
/clone="RP11-54C4"
BASE COUNT 50024 a 34373 g 48826 t

Query Match 49.7%; Score 684.2; DB 9; Length 170368;
Best Local Similarity 82.0%; Pred. No. 1.1e-191;
Matches 827; Conservative 0; Mismatches 148; Indels 34; Gaps 2;
QY 402 TAAGCAGCAAGCATTCAAGAGGTGACTTGGTGTCTGTGTTAAAGGCATTCAGTTTCATAG 461
DB 25774 TAAGCAGCAAGCATTCAAGAGTTATGTGTTGCTGTGTTAAAGGCATTCGGTTTATAAG 25715
QY 462 GGAGGCAGACATAAGAGTTGACAAAATTTGCACCCCTGACAATGTGATATAAAGAAAAA 521
DB 25714 GGAAGCAGAGCATAAAATTTGGAAAAATTTGCAGCCTGACATGCTAGAAAAAGAAAA 25655
QY 522 CCCATTTCTGAGGGAAATTCAGCTGCTGCAGAAATTTGCATATGTAATGAGAGCT 581
DB 25654 CCCATTTCTGAGGAAATTTGAAGCCAGCTGAAGAAATTTGCATAAGTAATGAGAGCC 25595
QY 582 GAATGTTAATCTCAAGACAATGGGAAAAATATCTCTGAGATGTCAGAGTCTTCACA 641
DB 25594 AAATGTTAATCCCCAAGACATAGGGAATGTCCTCGGAGCATGTCAGAGTCTTCACA 25535
QY 642 GCAGTCCATCAATCACTGGCCTGAGGCCTAGGAGAAATGTTTTGTGGAGCAGGCC 701
DB 25534 GCAGCCACTCCCATCACAGGCCAAAGGCTTTGGAGAAATGTTCTGTTGGCCAGGCC 25475
QY 702 AGGTCCTCTGCTGTGTCACGCTAGACACTTGTGTGCTCTGTGCCAGTTAATTCAGC 761
DB 25474 AGGTCCTCTGCTGTGTCGCTAGAGACTTGTGTGCTCTGTGCCAGTCCAC 25415
QY 762 -----TGTGGCTTCAGAGGTCAGGCCCAA 788
DB 25414 CATGACTAAAGGTGCCAAGTACAAATTCAGGTGTTGTTCTTCAGAGGTGGAGGCCCAA 25355
QY 789 GCCTTGGCAGCTTCCAAGTGGTGTGAGCCTGTGGGTGCAGAAAGTCAAGAAATTCAGGT 848
DB 25354 GCCTTGGCAGCTTCCACGTGGCATTGAGCCTGCGGTGCACAGAGTCAAGAAATTCAGGT 25295
QY 849 TTGGGAACCTCCANACAGATTTTCAGAGATATATGAAACCCCTGGATGCCAGGCAGAA 908
DB 25294 TTGGTAACCTCTGCCTAGATTTTCAGAGGTGTATGAAACCCCTGGATGCCAGGCAGAA 25235
QY 909 GTTTCGTGTAGGGTGGGCTCTCATGGAGAACCTCTGCAAGGTAGTACAAAGGAAA 968
DB 25234 GTCTGCTGAGGGTGGGCTCTCATGGAGAACCTCTGCTGAGGAGTGCAGAGGAAA 25175
QY 969 TGTGGTGGGAGCCCCACACAGAGTCCCCAGTGGGGTCCCTAGTAGAGCTGTGAG 1028
DB 25174 TGTGGGTTTGAAGCCCCACACAGAGTCCCTACTTGGGCACCTCTAGTAGGAGCTGTGAG 25115
QY 1029 AGAAGTCCACCATCTCTCAGACTCCAGAGGTTAGATCCACTGACAGCTTCAGCATGT 1088
DB 25114 AAAGCGGCCACACACTCCAGACCCAGAAATGTTAGATCCACTGACAACTTGCACGTGTGA 25055
QY 1089 GCTGAAAAATCCACAGACACTCAGTGCAGGCTGTGGAAGCAGCAGGAGTGTGCTGT 1148
DB 25054 GCTGAAAAAGCCACAGACACTCAACGCTAGCCTATGAAAGCAGCCAGGAGGAGGCTCT 24995

QY 1149 ACCCTACAAAAACCGTAGTGGCAGAGCTGACCAAGACCGTGGGAATCTACTCTTGCATTG 1208
DB 24994 ACCCTGAAAAGCCAGAGGGCATAGCTGCCAAGACCATGGGAACCCACCTTGCATCA 24935
QY 1209 TCATGACCTGGAGTCGAGACATGGAGTCAAAAGAGATCATTTGGAGCTTAAAGATTGA 1268
DB 24934 GTGTGACCTGCATATGAGACGTGGAGTCAAAAGAGATCATTTGGAGCTTAAAGACTTGA 24875
QY 1269 CTGCCCCACTGGATTTCGGACTTATATGGGGCCGTA-CCCCTTTGTTTGGCCAAATTT 1327
DB 24874 CTGTCTACTGGATTTCAGACTCGCATGGGGCTGTAGCCCTTTGTTTGGCCATGTC 24815
QY 1328 TTCCATTGGAACTGCCGTATTATACCAATGCTGTACCTCCATGTAT 1376
DB 24814 TCCCATTTAGATGGCTGTATTATCAATAGCTGTACCCCATGTAT 24766

RESULT 12
AC010312
LOCUS Homo sapiens chromosome 5 clone CTFB-62P13, complete sequence.
DEFINITION
AC010312
VERSION
AC010312.5 GI:15281196
KEYWORDS
HTG.
SOURCE
ORGANISM Homo sapiens.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE 1 (bases 1 to 67984)
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE Direct Submission
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 67984)
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Unpublished

REFERENCE 3 (bases 1 to 67984)
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE Direct Submission
JOURNAL Unpublished
REFERENCE 4 (bases 1 to 67984)
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE Direct Submission
JOURNAL Unpublished

REFERENCE 5 (bases 1 to 67984)
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE Direct Submission
JOURNAL Unpublished
COMMENT On Aug 23, 2001 this sequence version replaced gi:11079410.
Draft Sequence Produced by DOE Joint Genome Institute
www.jgi.doe.gov
Finishing Completed at Stanford Human Genome Center
www-shgc.stanford.edu

Quality: Phrap Quality >=40 99.7% of Sequence;
Estimated Total Number of Errors is 0.1.
STS Content:
SHGC-10855 G13672.
Location/Qualifiers
1. 67984
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="5"
/clone="CTB-62P13"

BASE COUNT 19168 a 13195 c 14662 g 20959 t
ORIGIN

Query Match 49.6%; Score 683.6; DB 9; Length 67984;
Best Local Similarity 81.9%; Pred. No. 1.5e-191;
Matches 827; Conservative 0; Mismatches 149; Indels 34; Gaps 2;

Qy	401	CTAAGCAGCAAGCATTCAGAGAGTGACTTGGGTGCTGTTAAAGGCATTCAGTTCTATAA	460
Ds	7119	CTAAGCAGCAAGCATTCAGAGAGTGATGTGGTGTCTGTTAAAGGCACCTCGGTTTATAA	7178
Qy	461	GGGAGGCAGACATTAAGAGTTCAGAAAATTTGCACCCCTGCACATGTGATAAAGAAAA	520
Ds	7179	GGGAGGCAGACATTAAGAGTTCAGAAAATTTGCACCCCTGCACATGTGATAAAGAAAA	7238
Qy	521	ACCCATTTTCTGAGGGGAAATTTCAAGCTGGCTGCAGAAAATTTGCATATGTAATGAGGAGC	580
Ds	7239	ACCCATTTTCTGAGGGGAAATTTCAAGCTGGCTGCAGAAAATTTGCATATGTAATGAGGAGC	7298
Qy	581	TGAATGTTAATCCTCAAGACAAATGGGAAAAATATCTCTCGACATGTGCAGAGTCTTCCAC	640
Ds	7299	CAAAATGTTAATCCTCAAGACAAATGGGAAAAATATCTCTCGACATGTGCAGAGTCTTCCAC	7358
Qy	641	AGCAGTCCATCAATCACTGCGCTGGAGCCCTAGGAGAAAATGTTTGTGGGACAGGCC	700
Ds	7359	AGCAGTCCATCAATCACTGCGCTGGAGCCCTAGGAGAAAATGTTTGTGGGACAGGCC	7418
Qy	701	CAGGTCCTCTGTGCTGTGCAGCCTAGAGACTTGGTGCCTGCTGCCAGTTAATTCAG	760
Ds	7419	CAGGTCCTCTGTGCTGTGCAGCCTAGAGACTTGGTGCCTGCTGCCAGTTAATTCAG	7478
Qy	761	C-----TGTTGGCTTCAGAGGGTGCAGGCCCA	787
Ds	7479	CCATGACTAAAAAGGTGCCAAAAGTACAATTCAGGTGTTTCTTCAGAAGGTGGAAGCCCA	7538
Qy	788	AGCCTTGGCAGCTTCCAAGTGGTGTGAGCCTGTGGGTGCGAAAGTCAAGATTGAGG	847
Ds	7539	AGCCTTGGCAGCTTCCAAGTGGTGTGAGCCTGTGGGTGCGAAAGTCAAGATTGAGG	7598
Qy	848	TTTGGAACTTCCAATCAGATTTTCAGAGATATATGGAACCCCTGGATGCCAGGCAGA	907
Ds	7599	TTTGGTAACTTCCCTAGATTTTCAGAGATATATGGAACCCCTGGATGCCAGGCAGA	7658
Qy	908	AGTTTGTGTAGGGTGGGTTCCTCATGTGAGAACTCTTCGAAGGGTAGTACAAAAGGAA	967
Ds	7659	AGTCTGTGCAAGGGTGGGCTCCTCATGGAGAACCTTCGTAGGCGAGTGCAGAAGGAA	7718
Qy	968	ATGTTGGTGGAGCCGCCACACAGAGTCCCGAGTGGGGCTCCATCTAGTAGAGCTGTGA	1027
Ds	7719	ATGTTGGTGGAGCCGCCACACAGAGTCCCGAGTGGGGCTCCATCTAGTAGAGCTGTGA	7778
Qy	1028	GAAGAAGTCCACCATCTCCAGACTCCAGAGGGTAGATCCACTGACAGCTTGCGAGCATG	1087
Ds	7779	GAAGAGCCACCATCTCCAGACTCCAGAGGGTAGATCCACTGACAGCTTGCGAGCATG	7838
Qy	1088	TGCCTGAAAAATCCACAGACACTCAGTGCCAGCCCTGTGAAAGCAGCAGGGATGGAGTCTG	1147
Ds	7839	AGCCTGAAAAAGCCACAGACACTCAACGCTAGCCCTATGAAAGCAGCAGGGAGGGCTC	7898
Qy	1148	TACCTAGAAAAACCGTAGTGCAGAGCTGACCAAGCCGTGGGATCTACCTCTTGCAT	1207
Ds	7899	TACCTAGAAAAACCGTAGTGCAGAGCTGACCAAGCCGTGGGATCTACCTCTTGCAT	7958
Qy	1208	GTCATGACCTGACGTGAGACATGGAGTCAAAAGAGATCATTTTGGAGCTTTAAGATTG	1267
Ds	7959	AGTGTGACCTGCATATGAGACGTGGAGTCAAAAGAGATCATTTTGGAGCTTTAAGATTG	8018
Qy	1268	ACTGCCCACTGGATTTTGGGCTATATATGAGGGCCCGTA-CGCCCTTGTGTTGGCCAAAT	1326
Ds	8019	ACTGTCTACTGGATTTTGGGCTATATATGAGGGCCCGTA-CGCCCTTGTGTTGGCCAAAT	8078
Qy	1327	TTTCCATTTGGAATCGCGTATTTACCAATGCGCTGTACCTTCATTTGAT	1376
Ds	8079	CTCCCATTTGGAATCGCGTATTTACCAATGCGCTGTACCTTCATTTGAT	8128
RESULT 13			
AC091005/c			
LOCUS	AC091005	168502 bp	DNA linear PRI 11-DEC-2001

DEFINITION
AC091005
VERSION
AC091005.9 GI:17488655
KEYWORDS
HTG.
SOURCE
Homo sapiens.
ORGANISM
Homo sapiens

REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS

Homo sapiens chromosome 15, clone RP11-1008C21, complete sequence.
1 (bases 1 to 168502)
Homo sapiens chromosome 15, clone RP11-1008C21
Unpublished
2 (bases 1 to 168502)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,S.,
Barna,N., Bastien,V., Boguslavskiy,L., Bouckghalter,B., Brown,A.,
Camarata,J., Campopiano,A., Chang,J., Choepel,Y., Colangelo,M.,
Collins,S., Collymore,A., Cooke,P., DeArellano,K., Dewar,K.,
Diaz,J.S., Dodge,S., Faro,S., Ferreira,P., FitzHugh,W., Gage,D.,
Galagan,J., Gardyna,S., Glnde,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Hagos,B., Heaford,A., Horton,L., Hulme,W.,
Iliev,I., Johnson,R., Jones,C., Karatas,A., LaRocque,K.,
Lamazares,R., Landers,T., Lehoczy,J., Levine,R., Liu,G.,
MacLean,C., Macdonald,P., Marquis,N., Matthews,C., McCarthy,M.,
McEwan,P., McKernan,K., McPheeters,R., Meldrim,J., Meneus,L.,
Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Norbu,C.,
Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J.,
Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C.,
Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J.,
Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R., Seaman,S.,
Severy,P., Sougnez,C., Spencer,B., Stange-Thomann,N.,
Stojanovic,N., Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S.,
Theodore,J., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.

TITLE
JOURNAL
REFERENCE
AUTHORS

Direct Submission
Submitted (24-MAR-2001) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 168502)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
Anderson,S., Barna,N., Bastien,V., Boguslavskiy,L., Bouckghalter,B.,
Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,A.,
Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
Glnde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
Jones,C., Kanat,A., Karatas,A., Kells,C., LaRocque,K.,
Lamazares,R., Landers,T., Lehoczy,J., Levine,R., Liu,G.,
MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
McCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Meldrim,J.,
Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C.,
Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D.,
Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.

TITLE
JOURNAL
REFERENCE
AUTHORS

Direct Submission
Submitted (01-SEP-2001) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
4 (bases 1 to 168502)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
Anderson,S., Barna,N., Bastien,V., Boguslavskiy,L., Bouckghalter,B.,
Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,A.,
Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
Glnde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
Jones,C., Kanat,A., Karatas,A., Kells,C., LaRocque,K.,

Lamazares,R., Landers,T., Lehoczkzy,J., Levine,R., Liu,G.,
 MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
 McCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Meldrim,J.,
 Menes,L., Mihova,T., Mienga,V., Murphy,T., Naylor,J., Nguyen,C.,
 Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D.,
 Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
 Raymond,C., Retta,R., Riback,M., Riley,R., Rise,C., Rogov,P.,
 Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
 Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
 Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
 Topham,K., Travers,M., Travis,N., Trigliio,J., Vassiliev,H.,
 Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye.W.J., Young,G.,
 Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: <http://www-seq.wi.mit.edu>

Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information

Center project name: LI0835

Center clone name: 1008_C_21

FEATURES

source

Location/Qualifiers

1..168502
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="15"
 /map="15"
 /clone="Rp11-1008C21"
 /clone_lib="RPC1-11 Human Male BAC"
 140..191
 /rpt_family="L2"
 194..218
 /rpt_family="(CAA)n"
 complement(222..381)
 /rpt_family="FRAM"
 384..464
 /rpt_family="L2"
 412..417
 /note="<30 qual SINGL region"
 667..672
 /note="<30 qual SINGL region"
 799..803
 /note="<30 qual SINGL region"
 1130..1158
 /rpt_family="(CA)n"
 complement(1886..1985)
 /rpt_family="MIR"
 complement(2627..2637)
 /note="<30 qual SINGL region"
 3489..3530
 /rpt_family="(A)n"
 4319..4348
 /rpt_family="AT-rich"
 4775..4903
 /rpt_family="MER85"
 5137..5175
 /rpt_family="AT-rich"
 5198..5557
 /rpt_family="MLT2F"
 5558..5595
 /rpt_family="(CA)n"
 5596..5624
 /rpt_family="MLT2F"
 complement(6251..6387)
 /rpt_family="MIR"

unsure 6377..6382
 /note="<30 qual SINGL region"
 complement(6414..6621)
 /rpt_family="MIR"
 6744..6963
 /rpt_family="L2"
 complement(7718..8106)
 /rpt_family="MLT1B"
 8610..8732
 /rpt_family="L2"
 complement(9591..9711)
 /rpt_family="MIR"
 complement(10164..10479)
 /rpt_family="AluJo"
 complement(11119..11206)
 /rpt_family="MIR"
 13093..13155
 /rpt_family="(CGTG)n"
 13446..13751
 /rpt_family="AluX"
 14102..14669
 /rpt_family="MLT1E2"
 14671..14973
 /rpt_family="AluJb"
 15003..15117
 /rpt_family="(GAAA)n"
 complement(15227..15559)
 /rpt_family="AluJb"
 15738..16076
 /rpt_family="MLT1L"
 17512..17576
 /rpt_family="MLT1L"
 17763..17980
 /rpt_family="MERS8A"
 18048..18230
 /rpt_family="MERSA"
 18427..18562
 /rpt_family="MIR3"
 complement(18660..18828)
 /rpt_family="MERSB"
 complement(18937..19116)
 /rpt_family="MERSA"
 complement(19619..20141)
 /rpt_family="L2"
 complement(20145..20670)
 /rpt_family="MER69"
 complement(20669..21182)
 /rpt_family="MER69"
 complement(21183..21315)
 /rpt_family="L2"
 complement(21860..22086)
 /rpt_family="MERS8A"
 complement(22428..22915)

Query Match 49.6%; Score 683.2; DB 9; Length 168502;
 Best Local Similarity 80.4%; Pred. No. 2.1e-191;
 Matches 852; Conservative 0; Mismatches 173; Indels 35; Gaps 3;

QY 351 GAATATTTGGCTAATGAGGAACAAATTTTGGCTAATAGGAAGAAATTTCTAAGCAGCA 410
 Db 45723 GAACTTTGAACCTTGAGAGAGATGATTTAGGGTATCTGGCAGAAGAAATTTCTAAGCAGCA 45664
 QY 411 AAGCATTCAGAGAGTGTGGTGTCTTAAGAGCATTTCAGTTTCATAGGAGGAGCACA 470
 Db 45663 AAGCATTTTCAGATGTGATGGTGTCTGTTAAGGCAATTCGGTTTTATAGGAAGCAGCA 45604
 QY 471 GCATAAGAGTTTCAGAAAATTTGCACCCCTGACAAATGTGATAAAAAAGAAAAACCCATTTTC 530
 Db 45603 GCATGAAAGCTTTGGAAAAATTCAGCCTGACAAATGTGATAGAAAAAGAAAACTTCA-TTTC 45545
 QY 531 TGAGGGGAAATTCAGCTGGCTGCAGAAATTTGCATATATGATGAGGAGCTGAATGTTAA 590
 Db 45544 TGAGGAGAAATTCAGCTGGCTTCATAAATTTGCATAAGCAATGAGGAGCAATGTTTCA 45485


```
Qy 591 TCCTCAAGCAATGGGAAAAATATCTCTCGACATGTCTCAGAGGTCTTCACAGCAGTCCAT 650
Dy 4584 TCCCAAGACAATGGGAAAAATGTCTCCAGGCATGTCTCAGAGGTCTTCATAACAGCCCT 45425
Qy 651 CAATCACTGGCCTGGAGCCCTAGGAGAAATGGTTTGTGGAGAGCCAGGTCCTCT 710
Dy 45424 GCCATCAGAGCCCTGGAGCCCTAGGAGAAATGGTTTGTGGAGAGCCAGGTCCTCC 45365
Qy 711 GTGCTGTGTGACGCTAGAGACTTGGTCCCTGTGTCCAGTTAATTCAG----- 760
Dy 45364 ATGCTGTGTGACGCTAGGACTTGTTCCTGTGATCCAGCTGCTCCAGCATGGCTGA 45305
Qy 761 -----CTGTGGCTTCAGAGGGTGCAGCCCAAGCCCAAGCCCTGGCA 45305
Dy 45304 AAGGGCCCAACATAGAGCTGGACTGTGCTTCAGAGGTGGAAGCCCAAGCCCTGACA 45245
Qy 798 GCTTCCAAAGTGGTGTGAGCCTGTGGTGCAGAGAGTCAAGATTTGAGGTTTGGGAACC 857
Dy 45244 GCTTCCAGCTGGTGTGAGCCTTACAGGTGCACAGAGTCAAGAACTGGGGTTTGGGAACC 45185
Qy 858 TCCAAATCAGATTTCAGAGAACTATATGAAACCCCTGGATGCCAGGCAGAAATTTGCTGT 917
Dy 45184 TCCGCTAGATTTAGAGATGTATAGAAACACCTGGATGCCAGGCAGAAATTTGCTGT 45125
Qy 918 AGGGTGGGGTCTCATGGAGAACTCTGCAAGGGTAGTACAAAGGAGAAATGTTGGGTG 977
Dy 45124 AGGGTGGGGATTCATGGAGAACTCTACTAGGCAGCATGGAAGGAGAAATGTTGGGTG 45065
Qy 978 GGAGCCCCACACAGAGTCCCCAGTGGGCTCCATCTAGTAGAGCTGTGAGAGAGTCC 1037
Dy 45064 AGATCCCCACACAGAGTCCCTCTGGGGCTGCTCTAGTGAGAGCTGTGAGAGAGGGCC 45005
Qy 1038 ACCATCTCCAGACTCCAGAGGGTAGATCCACTGACAGCTTGACAGCATGTGCCTGAAAA 1097
Dy 45004 ACTGCTCTCCAGACCCAGATGGTAGATCCACCAGCAGCTTGACCAATTTGCCTGGAAA 44945
Qy 1098 ATCCACAGACACTAGTCCCGCCTGTGAAAGCAGCAGGAGTGTGACCTTACCTACAA 1157
Dy 44944 AAGGCAGACACTCAATCCAGCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 44885
Qy 1158 AACCTAGTGCAGAGCTGACAGACCGTGGATCTACCTCTGATGATGATGATGATGATG 1217
Dy 44884 AACCAAGGGTGGAGTGGCCAGACCATGGAGCCACCTCTTGCATGATGATGATGATGAT 44825
Qy 1218 GGAGTGCAGATGAGTCAAGAGATCAATTTTGAGCTTTTAAAGATTTGAGTGCCTCCAC 1277
Dy 44824 GGATGTGAGACTGGAGTCAAGAGATCAATTTTGAGCTTTTAAAGATTTGAGTGCCTCCAC 44765
Qy 1278 TGGATTTCCGAGCTTATATGGGGCCGT-ACCCCTTTGTTTGGCCAAATTTTTCATTTG 1336
Dy 44764 TGGATTTGGGACTTGCATGGGCCCTGTAAACCCCTTTGTTAGGCCAAATTTTCCCATTTG 44705
Qy 1337 GAAGTCCGCTATTTACCAATGCTGTACCTCCATTTGAT 1376
Dy 44704 GAATGGCTGATTTACCCAAATACCTGTACTGCTCCATTTGAT 44665

RESULT 14
AC079614 AC079614 131215 bp DNA linear PRI 09-JAN-2002
DEFINITION Homo sapiens BAC clone RP11-42414 from 2, complete sequence.
ACCESSION AC079614
VERSION AC079614.7 GI:15431230
KEYWORDS HTG.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 131215)
REFERENCE Sulston, J.E. and Waterston, R.
AUTHORS Toward a complete human genome sequence
TITLE Genome Res. 8 (11), 1097-1108 (1998)
JOURNAL
```

```
99063792
9847074
2 (bases 1 to 131215)
Nguyen, C., Abbott, A., and Elliott, G.
The sequence of Homo sapiens BAC clone RP11-42414
Unpublished (2001)
3 (bases 1 to 131215)
Waterston, R.H.
Direct Submission
Submitted (04-SEP-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
4 (bases 1 to 131215)
Waterston, R.H.
Direct Submission
Submitted (05-SEP-2001) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
5 (bases 1 to 131215)
Waterston, R.
Direct Submission
Submitted (09-JAN-2002) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
On Sep 5, 2001 this sequence version replaced gi:15209227.
----- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc
Contact: saplens@wustl.edu
----- Summary Statistics
----- Center project name: H_NH0424104

NOTICE: This sequence may not represent the entire insert of this
clone. It may be shorter because we only sequence overlapping
clone sections once, or longer because we provide a small overlap
between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded, sequenced with an alternate
chemistry, or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by sequence
from more than one subclone; and the assembly was confirmed by
restriction digest.

MAPPING INFORMATION:
Mapping information for this clone was provided by Dr. John D.
McPherson, Department of Genetics, Washington University, St. Louis
MO. For additional information about the map position of this
sequence, see http://genome.wustl.edu/gsc

SOURCE INFORMATION:
The RPCI-11 human BAC library was made from the blood of one male
donor, as described by Osoegawa, K., Woon, P.Y., Zhao, B., Frengen, E.,
Tateno, M., Catanese, J.J., and de Jong, P.J. (1998) An improved
approach for construction of bacterial artificial chromosome
libraries. Genomics 51:1-8. The clone may be obtained either from
Research Genetics, Inc. (http://www.resgen.com) or Pieter de Jong
and coworkers at the Roswell Park Cancer Institute
(http://bacpac.med.buffalo.edu)
VECTOR: pBACe3.6
NEIGHBORING SEQUENCE INFORMATION:
The clone sequenced to the left is RP11-22L19, 2000 bp overlap; the
clone sequenced to the right is RP11-364I15. Actual start of this
clone is at base position 128235 of RP11-22L19; actual end is at
base position 131215 of RP11-42414.

The sequence between 58007 and 58074 is covered only by a PCR
product of clone DNA; the sequence between 58180 and 58221 is
covered only by a PCR product of clone DNA. Data from AC093650 was
used to finish this clone, AC079614. Polymorphisms exist between
RP11-42414 and RP11-364I15.
```



```
OY 785 CCAAGCCCTGGCAGCTTCCCAAGTGGTGTAGACCTGTGGGTGCGAAAGAGTCAAGAATTG 844
|||||
Db 128705 TCAAGCCCTGGCAGCTTCCCAAGTGGTGTAGACCTGTGGGTGCGAAAGAGTCAAGAATTG 128764
OY 845 AGTTTGGGAACTTCCCAATCAGATTTCCAGAGATATATGGAACCCCTGGATGCCAGGC 904
|||||
Db 128765 AGTTTGGGAACTTCCCAATCAGATTTCCAGAGATATATGGAACCCCTGGATGCCAGGC 128824
OY 905 AGAAGTTTGTGTAGGGGTGGGCTTCATGGAGAACCTCTGCAAGGGTGTAGTACAAAAG 964
|||||
Db 128825 AGAAGTTTGTGTAGGGGTGGGCTTCATGGAGAACCTCTGCAAGGGTGTAGTACAAAAG 128884
OY 965 GAAATGTGGTGGGAGCCGCCACACAGAGTCCCAGTGGGGCTCCATCTAGTAGAGCTG 1024
|||||
Db 128885 GAAATGTGGTGGGAGCCGCCACACAGAGTCCCAGTGGGGCTCCATCTAGTAGAGCTG 128944
OY 1025 TGACGAAGAGTCCACCATCTCCAGAGTCCAGAGGTTAGATCCATCCACAGCTTGCAGC 1084
|||||
Db 128945 TGACGAAGAGTCCACCATCTCCAGAGTCCAGAGGTTAGATCCATCCACAGCTTGCAGC 129004
OY 1085 ATGTGCTGCTGAAAATCCAC--AGACACTCAGTGCAGCCTGTGAAAGCAGCAGGATGGA 1142
|||||
Db 129005 ATGTGCTGCTGAAAATCCAC--AGACACTCAGTGCAGCCTGTGAAAGCAGCAGGATGGA 129064
OY 1143 GTCTGTACCTTACAAAACCGTAGTGGCAGAGCTGACCAAGACCGTGGGAATCTACCTCTT 1202
|||||
Db 129065 GCGTGTGACATTTCAAAGCCACAGCGGAGAGCTGCTCCAGAGTGGTGGGAGTGGGATG 129124
OY 1203 GCATTGTGACCTGTCAGCTGACATGGAGTCAAAAGAGATCATTTTGGAGCTTTAAG 1262
|||||
Db 129125 GCATTGTGACCTGTCAGCTGACATGGAGTCAAAAGAGATCATTTTGGAGCTTTAAG 129184
OY 1263 ATTTGACTGCCCACTGGATTTCCGACTATATGCGGCCCT-ACCCTTTGTTTGGCC 1321
|||||
Db 129185 ATTTGACTGCCCACTGGATTTCCGACTATATGCGGCCCT-ACCCTTTGTTTGGCC 129244
OY 1322 AATTTTTCATTTGGAACCTGCCGATTTTACCCCAATGCC 1361
|||||
Db 129245 AACATCTCCCATTTGGAATGGCTGCTATTTACCCAGTGCCT 129284
```

RESULT 15

AC012038/c

LOCUS

DEFINITION Homo sapiens 12 BAC RP11-547C5 (Roswell Park Cancer Institute Human BAC Library) complete sequence.

AC012038

AC012038.19 GI:12000459

VERSION

KEYWORDS

SOURCE

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.

1 (bases 1 to 73390)

Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,

Alsbrooks,S.L., Amaralunga,H.C., Are,J.R., Banks,T., Barbara,J.,

Benton,J., Blmage,K., Blankenburg,K., Bonnin,D., Bouck,J.,

Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,

Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,

Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,

Chen,Z., Chiu,D., Chowdhry,I., Christopoulos,C., Cleveland,C.D.,

Cox,C., Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,

Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,

Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,

Dugan-Rocha,S., Durbun,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,

Elhaj,C., Emerling,S., Escotto,M., Falls,T., Ferraguto,D.,

Flagg,N., Ford,J., Foster,P., Frantz,P., Gabisi,A., Gao,J.,

Garcia,A., Garner,T., Garza,N., Gill,R., Gorrell,J.H., Guevara,W.,

Gunaratne,P., Hale,S., Hamilton,K., Han,J., Harris,C., Harris,K.,

Hart,M., Havlak,P., Hawes,A., Hernandez,J., Hernandez,O.,

Hodgson,A., Hoques,M., Holloway,C., Hollins,B., Homsif,F.,

Howard,S., Huber,J., Hulyk,S., Hume,J., Ioshikhes,I., Jackson,L.E.,

Jacobson,B., Jla,Y., Johnson,R., Jolivet,S., Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lee,E., Lewis,L.C., Lewis,L., Li,J., Li,Z., Lichtarge,O., Lileu,C., Lileu,J., Lileu,W., Loulseged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J., Maheshwari,M., Mapua,P., Marondel,I., Martin,R., Martindale,A., Martinez,E., Massey,E., Mawhney,E., McLeod,M.P., Meador,M., Mei,G., Merscher,S., Metzger,M., Miller,A., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Montgomery,K.T., Morgan,M., Morris,S., Moser,M., Neal,D., Nelson,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S., Ogih,M., Okwunonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L., Peters,L., Pickens,R., Prljubokan,I., Rofle,M., Rulz,S., Savary,G., Scherer,S., Scott,G., Shen,H., Shalm,C., Shostakar,N., Sisson,I., Sodergren,E., Sonalke,T., Sparks,A., Stanley,H., Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalona,D., Vinson,R., Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S., Williams,G., Williamson,A., Wleczkyk,R., Wooden,S., Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorilla,S., Kucherlapati,R. and Glibbs,R.

Direct Submission

Unpublished

2 (bases 1 to 73390)

Worley,K.C.

Direct Submission

Submitted (19-OCT-1999)

Human Genome Sequencing Center, Department

Of Molecular and Human Genetics, Baylor College of Medicine, One

Baylor Plaza, Houston, TX 77030, USA

3 (bases 1 to 73390)

Worley,K.C.

Direct Submission

Submitted (01-JAN-2001)

Human Genome Sequencing Center, Department

Of Molecular and Human Genetics, Baylor College of Medicine, One

Baylor Plaza, Houston, TX 77030, USA

On Jan 1, 2001 this sequence version replaced gl:11138159.

INFORMATION: <http://www.hgsc.bcm.tmc.edu/> or email

gc-help@bcm.tmc.edu

CLONE LENGTH: This sequence does not necessarily represent the entire insert of this clone. Overlapping regions of clones are only sequenced and submitted once, so the sequence for the remainder of the insert may be found in the record for the adjacent clones. Overlapping clones are noted at the beginning and end of the Features listing.

ANNOTATION OF FEATURES:

STSs are identified using ePCR (Genome Res. 7:541-550) searches of a local database that includes entries from dbSTS, GDB, and local mapping efforts.

Repeats are identified using RepeatMasker (A. Smit and P. Green, unpublished.) for Human and Mouse sequences.

Genes and Region of sequence similarity are identified by BLAST (Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the EST and cDNA sequences. Genes demonstrate at least two exons flanked by consensus splice sites that maintained sequence continuity across the splice junctions. Sequences that are not identical matches are annotated as similar.

SEQUENCING READ COVERAGE: Sequencing is completed to a minimum standard of double strand coverage with a minimum of 2 clones and 2 reads with no ambiguities or 2 chemistries with a minimum of 2 clones and 3 reads with no ambiguities. If the sequence quality for a region does not meet this standard, it will be indicated in the annotation as Low Coverage.

QUALITY OF INDIVIDUAL BASES: This sequence meets stringent quality standards - estimated error rate less than 1 per 10,000 bases. Reports of lowest quality individual bases and measures of base quality are listed below. Description of the metrics can be found at URL: <http://gc.bcm.tmc.edu:8088/quality.info/genbank.annotation>.

html.

QUALSTAT-REPORT-----

----- Summary Statistics -----
Contig length: 73390
Phrap values in estimate: 73148
Average error rate (BCM-Phrap estimate): 4.24674e-07
Fraction of Phrap values less than 40 : 0.000546837
Number of consensus changing edits: 6
Number of N's in consensus : 0

----- Consensus changing edits -----
Position Original-Context Edited-Context
2854 ttgggcagg(g)acacaaatcc
40628 tccctgtgt(n)ncaaagggtg
40629 cctgtgtg(n)caagggtg
40678 ctgacacag(n)actcctgagn
40688 nactcctgag(n)tacccttcac
58519 tgaatttcac(n)cagaatcagg

----- Distribution of Quality < 40 Bases -----

1000	5	10	15	20	25	30	35	40
900								
800								
700								
# bases								
600								
500	*							
400	*							
300	*							
200	*	*	*	*	*	*	*	*
100	*	*	*	*	*	*	*	*
0	*	*	*	*	*	*	*	*

Phrap value Range

Version: 1.01 qxf0.
FEATURES Location/Qualifiers
source 1. .73390
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="12"
/clone="RP11-547C5"

repeat_region	19. .106
repeat_region	/rpt_family="MIR"
repeat_region	complement(826. .1067)
repeat_region	/rpt_family="MSTA"
repeat_region	complement(1070. .2183)
repeat_region	/rpt_family="MSTA-internal"
repeat_region	complement(2211. .2493)
repeat_region	/rpt_family="MSTA-internal"
repeat_region	complement(2494. .2875)
repeat_region	/rpt_family="MSTA"
repeat_region	complement(3718. .4776)
repeat_region	/rpt_family="L2"
repeat_region	5066. .5137
repeat_region	/rpt_family="MER5A"
repeat_region	complement(5082. .5228)
repeat_region	/rpt_family="MER5A"
repeat_region	5802. .6088
repeat_region	/rpt_family="MLT2"
repeat_region	6169. .6327
repeat_region	/rpt_family="MLT2H"
repeat_region	complement(6751. .6903)
repeat_region	/rpt_family="MLT2H"
repeat_region	6930. .7009

Query Match 49.4%; Score 680; DB 9; Length 73390;
Best Local Similarity 80.2%; Pred.No. 1.8e-190;
Matches 850; Conservative 0; Mismatches 175; Indels 35; Gaps 3;

QY	351	GAATATTTGGCTAATGAGGAAGCAAAATTTTGGCTAATAAGGAAGAAATTTCTTAAGCAGCA	410
DB	65115	GAACTTTGAACCTTGAGAGAGATGATTAGGGTATCTGGCGGAAGAAATTTCTTAAGAGCA	65056
QY	411	AAGCATTCAGAGGTGACCTTGGGTGCTGTTAAAGGCAATTCAGTTTCATAGGAGGACAGA	470
DB	65055	AAGCATTCAGAGGTGACCTTGGGTGCTGTTATAGGCATTAAGTTTATAGGGAAGGAGA	64996
QY	471	GCATAGAGTTCAGAAAATTTGCACCCTGCACATGTGATAAAGAAAAGAAACCCATTTTC	530
DB	64995	GCATAAAGTTTCAGAAAATTTGCACCCTGCACATGTGGTAGAAAAGAAACCCATTTTC	64936
QY	531	TGAGGAAAATTCAGCTGGCTGCAGAAATTTGCATATGTAATAGGAGCTGAATGTTAA	590
DB	64935	TGAGGAAAATTCAGCCAGCTGCAGAAATTTGCATATGTAATAGGAGCCCAATATTTAA	64876
QY	591	TCCTCAAGACAATGGGAAAATATCTCTGGACATGTACAGAGTCTTCACAGAGTCCAT	650
DB	64875	TTGCCAAGACAATGGGAAAATGTCTCCAGGCGATGTACAGAGTCTTCACAGCAACCCCT	64816
QY	651	CAAACTCACTGGCTGGAGCCCTAGGAGAAAATGGTTTGTGGACAGGCCCGGTCCT	710
DB	64815	CCATCAAGGCTCAGAGACCTTAGGAGAAATGGTTTTCATGGCGCCAGCCCGGTCCT	64756
QY	711	GTGCTGTGTGCAGCTAGAGACTTGTGGCCCTGTGTCCAGCTTAATTTCA- - - - -	759
DB	64755	GTGCTGTGTGCAGCTAGAGACTTGTGGCCCTGTGTCCAGCTGTTCAGACCATGGGTAA	64696
QY	760	-----GCTGTGGCTTCAGAGGTGCAAGCCCAAGCCCTTTGGCA	797
DB	64695	AAGGGGCAATTAGAGCTCAGGCCATGGCTTCAGAGGTGCAAGCTCCAAAGCTTTGCA	64636
QY	798	GCTTCCAAGTGGTGTGAGCCCTGGGTGCAAGAGTCAAGAAATAGTTTTGGGAACC	857
DB	64635	GCTTCCAAGTGGTGCAGAGCCCTGCAAGTGCACAGAGTCAAGAAATAGTTTTGGGAACC	64576
QY	858	TCCAATCAGATTTCAGAGATATATGGAAACCCCTGGATGCCCGCAGCAGAACTTTGCTCT	917
DB	64575	TCTGCTCAGATTTCAGAGATATATGGAAACACCTGGATGCCCGCAGCAGAACTTTGCTCT	64516
QY	918	AGGGGTGGGTCTCATGGAGAACCTCTGCAAGAGGTAGTACAAAAGGGAAATTTGGGTG	977
DB	64515	AGGGGTGGGTCTCATGGAGAACCTCTGCTGGGGCAGTGTGGAAGGAAATTTGGGTG	64456

THIS PAGE BLANK (USPTO)